Identifying Oxide Defects in Semiconductor Devices Using Carrier Capture and Emission Rate Calculations

W. Goes*, Y. Wimmer†, A.-M. El-Sayed°,*, G. Rzepa*, M. Jech*, A. L. Shluger°, T. Grasser*
*Institute for Microelectronics, TU Wien, A-1040 Vienna, Austria
†Christian Doppler Laboratory for Reliability Issues in Microelectronics
°Department of Physics and Astronomy and London Centre for Nanotechnology, University College London, Gower Street, London WC1E 6BT, United Kingdom
(Dated: December 4, 2017)

It is well-established that oxide defects adversely affect functionality and reliability of a wide range of microelectronic devices. In semiconductor-insulator systems, insulator defects can capture or emit charge carriers from/to the semiconductor. These defects feature several stable configurations, which may have profound implications for the rates of the charge capture and emission processes. Recently, these complex capture/emission events have been investigated experimentally in considerable detail in Si/SiO_2 devices, but their theoretical understanding still remains vague. In this paper we discuss in detail how the capture/emission processes can be simulated using the theoretical methods developed for calculating rates of charge transfer reactions between molecules and in electro-chemistry. By employing this theoretical framework we link the atomistic defect configurations to known trapping model parameters (e.g. trap levels) as well as measured capture/emission times in Si/SiO_2 devices. Using density functional theory (DFT) calculations, we investigate possible atomistic configurations for various defects in amorphous (a)-SiO_2 implicated in being involved in the degradation of microelectronic devices. These include the oxygen vacancy and hydrogen bridge as well as the recently proposed hydroxyl E' center. In order to capture the effects of statistical defect-to-defect variations that are inevitably present in amorphous insulators, we analyze a large ensemble of defects both experimentally and theoretically. This large-scale investigation allows us to prioritize the candidates from our defect list based on their trap parameter distributions. For example, we can rule out the E' center as a possible candidate. In addition, we establish realistic ranges for the trap parameters, which are useful for model calibration and increase the credibility of simulation results by avoiding artificial solutions. Furthermore, we address the effect of nuclear tunneling, which is involved according to the theory of charge transfer reactions. Based on our DFT results, we demonstrate the impact of nuclear tunneling on the capture/emission process, including their temperature and field dependence, and also give estimates for this effect in Si/SiO_2 devices.

CONTENTS

I. Introduction 1
II. Findings from TDDS Studies 3
III. Derivation of the Defect Model 5
III.1. Adiabatic Charge Transfer Reactions 5
III.2. Nonadiabatic Charge Transfer Reactions 8
III.3. Fermi's Golden Rule 9
III.4. Franck-Condon Principle 9
III.5. Thermally Activated Transitions 11
III.6. Electronic Matrix Element 12
III.7. Evaluation of the Franck-Condon Factors 12
III.8. Configuration Coordinate Diagram 14
III.8.1. Electron Trap 14
III.8.2. Hole Trap 17
III.9. Interactions with the Conduction/Valence Band 19
IV. Stochastic Dynamics of Electron Transfer 21
V. Defect Candidates 23
V.1. Details of the DFT Calculations 24
V.2. Oxygen Vacancy 24
V.3. Hydrogen Bridge 25
V.4. Hydroxyl E' Center 26
VI. Parameter Extraction from DFT 27
VI.1. Trap levels 27
VI.2. Impact of Quantum Effects on the Temperature Dependence 31
VI.2.1. High-Temperature Limit 34
VI.3. Impact of Quantum Effects on the Gate Bias Dependence 36
VI.4. Worst-Case Scenarios for the Classical Approximation 37
VII. Conclusions 39
VIII. Acknowledgements 40
Appendix 40
VIII.1. Hole Trap in the Hole Picture 40
VIII.2. Wenzel-Kramers-Brillouin Method 41
References 42

I. INTRODUCTION

Metaloxidesemiconductor transistors are used in a wide range of electronic devices underpinning 'internet

of things' and most technologies and therefore have become an indispensable part of our daily lives. Their technological success has been the result of intensive R&D efforts, which have continued over decades and led to substantial changes in technologies used for producing these devices. Nevertheless, the technologically most relevant field-effect transistors (FETs) are still based on semiconductor-insulator interface systems which are required for the field-effect and thus the functionality of the transistors. While idealized materials can be assumed for the basic understanding of their functional principles, real devices contain defects, particularly in the amorphous insulator and its interface to the semiconducting channel. For instance, P_b centers, three-coordinated Si atoms with a dangling bond, have been detected at $\mathrm{Si/SiO_2}$ interfaces by electron spin resonance (ESR) spectroscopy [1, 2]. In $\mathrm{SiO_2}$, this measurement method revealed the existence of several types of so called E' centers [1, 2], which originate from Si dangling bonds. Among them, the most prominent are variants of the E'_γ center usually associated with oxygen deficiency. In addition, the 74 and 10.4 G doublet center, hydrogenated variants of the E' center, have been discovered in $\mathrm{SiO_2}$ [3]. The existence of these ESR-active centers has also been confirmed by spin-dependent recombination experiments (SDR) [4]. Furthermore, investigations based on spin dependent tunneling (SDT) have revealed the existence of K_n centers in nitrided oxides [5].

In numerous electrical measurements, defects have been demonstrated to be electrically active, meaning that they can capture and emit charge carriers from the substrate as well as the gate. These defects, commonly referred to as charge traps, can strongly affect the device characteristics or even impair the functionality of the transistors. For instance, they were suggested to be involved in time-dependent dielectric breakdown, causing device failure [6]. Other prominent and serious issues, such as stress-induced leakage currents (SILC) [7–9], random-telegraph noise (RTN) in the drain current [10], bias temperature instability (BTI) [11]. These phenomena have not only been observed in conventional Si-based transistors but also in other important technologies: In SiC and GaN power devices, charge capture and emission were found to occur in such large concentrations that they dominate the device functionality [12, 13]. Furthermore, the introduction of novel two-dimensional materials is expected to provide high device performance in terms of switching speed and power consumption but has already been shown to suffer from charge trapping too. As such, charge trapping appears to be a widespread issue among all transistor technologies [14, 15].

For decades, these charge trapping phenomena have been investigated by different experimental techniques: For instance, RTN measurements and time-dependent defect spectroscopy (TDDS) can be used to detect single charge trapping events of selected defects and provide insight into the mechanisms behind charge trapping. Other measurement techniques are $I_\mathrm{d}(V_\mathrm{g})$ measurements [16–19], charge pumping [20, 21], measure-stress-measure [22–25], and on-the-fly measurements[26, 27] among others. However, all of them have initially been designed for large-area devices, in which a multitude of charge trapping events superimpose and thus single defects cannot be assessed experimentally.

In all these measurement techniques, charge trapping essentially occurs for two distinct operation modes of the transistors: In one mode, the gate bias is kept at a constant level and the charge capture and emission events occur stochastically distributed in time and give rise to drain current noise for instance. As such, an RTN signal corresponds to the equilibrium response for this operation mode with regard to charge trapping. In the other mode, the transistor is alternatingly operated at a high and a low gate bias level, where the former typically accelerates capture events while the latter accelerates emission. The acceleration due to the applied bias has motivated the term 'stimulated charge trapping' for this operation mode and will be used throughout this paper.

In order to better describe these phenomena theoretically, several models have been put forward. In those models, the charge trapping process has often been treated using phenomenological approaches [28] for calculating the capture and emission rates. More sophisticated models [29, 30] already accounted for the tunneling of charge carriers through the energy barrier from the substrate into the trap or vice versa. McWorther *et al.* [31–33] incorporated the tunneling effect into the Shockley-Read-Hall (SRH) theory in order to account for the broadly distributed capture/emission (C/E) times via a different trap depth. Kirton *et al.* [10] observed a strong temperature dependence of the C/E time constants in the measured drain current noise. In order to explain this, he also gave the capture/emission (C/E) process a first microscopical interpretation within the framework of the nonradiative multi-phonon (NMP) theory. This finding implied that the established SRH model is an oversimplifying description for C/E processes and is not applicable to charge trapping in oxide defects [34].

Even though some variants of the NMP theory have been applied in several studies [8, 35–38], their physical formulations often rely on simplifying assumptions. Most of them are based on a model Hamiltonian in which the interactions between the electrons and the phonons are described by a single term. As an approximation, this term is expanded in a series of the atomic coordinates $\mathbf{R}$ and assumed to be dominated by the first or second order term, referred to as 'linear' or 'quadratic' electron-phonon coupling, respectively. Although this approach allows for closed-form expressions of the nonradiative multi-phonon transitions, it does not capture the full complexity of actual charge capture or emission processes.

Similar processes have been thoroughly investigated under the term "charge transfer" (CT) reactions in the context of chemistry [39]. These reactions are well described by a variety of theories, which have been developed at various levels of sophistication. The most popular formulation was proposed by Marcus, who was

awarded the Nobel prize in chemistry for his pioneering work in 1992. Meanwhile, these theories have become well-established in many fields of chemistry and biology and are routinely used to describe processes, such as photosynthesis, corrosion, and chemiluminescence. Even though these theories led to a significant progress in the understanding of molecular conduction in nanoelectronics [40, 41], they have remained virtually unrecognized in the microelectronics community. In principle, the theory of CT reactions already lays the foundation for modeling the actual C/E processes involved in charge trapping. Defect model must still be adapted to the case of microelectronic transistors where semiconductor-insulator material systems are encountered. This requires consideration of all interactions with the band states in the semiconductor substrate and the gate contact, as it has already been done for molecular conduction [40, 41]. Furthermore, such a model must also account for the presence of additional defect states as suggested by various measurements [42, 43].

In order to gain an atomistic insight into charge trapping, the defects in the dielectrics were also investigated theoretically in numerous density functional theory (DFT) studies [44]. For instance, the P_b center has been intensively examined [45] and its reactions with atomic and molecular hydrogen considered in [46, 47]. Furthermore, the positively charged oxygen vacancy in SiO_2 was found to be stable in two configurations where one of them could be related to the E' center [48–50]. The property of having two stable configurations is also referred to as 'bi-stability' and maybe linked to some unexpected behavior of defects in microelectronic transistors. As such, it was the subject of intensive discussions but has been confirmed by several independent groups [49–55]. Looking from a device perspective, it has been shown that the oxygen vacancy easily reacts with a hydrogen and forms a defect called hydrogen bridge [7, 56], which has a defect level within the SiO_2 bandgap. In a number of publications [44, 57–59], several other defects have been found to introduce trap levels within the HfO_2 bandgap. Furthermore, possible bi-stable defect configurations were proposed by Joeng *et al.* [60] for nitrided oxides. In all those studies, however, the results of DFT calculations were not used to predict the electron capture and emission rates. In this paper we attempt to provide a link between the electronic properties of defects in the oxide and electron C/E rates measured experimentally.

To achieve that, we first present a rigorous derivation of a defect model to describe the various aspects of charge trapping and to link its parameters to atomistic simulations. To provide the experimental foundations, the findings from TDDS studies will be summarized and their implications for the defect model discussed in Sec. II. Next, the transition rates of C/E processes will be formulated based on the theory developed for CT reactions (in Sec. III). The resulting rate expressions will then be incorporated in the defect model suggested by TDDS experiments. Sec. V is devoted to defects which have been associated with charge trapping phenomena and investigated by DFT. In Sec. VI, their potential energy surfaces will be studied and related to the parameters of the defect model.

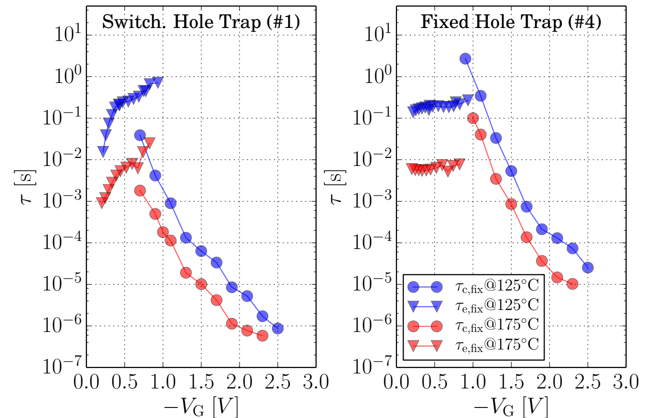


FIG. 1. Measured capture (circles) and emission (triangles) times of a switching (left) and a fixed (right) oxide hole trap. Both trap types show capture and emission times with a clear temperature activation and have a pronounced gate bias dependence in their capture times. However, they differ in the gate bias dependence of their emission times: While hole emission remains unaffected by the gate bias in the case of a fixed oxide hole trap, it is sensitive to bias variations for a switching hole trap.

II. FINDINGS FROM TDDS STUDIES

Recently a new measurement technique, termed TDDS, has shed additional light on charge trapping in microelectronic transistors [61, 62]. It makes use of the fact that single charge capture/emission events in small-area devices can be resolved as discrete steps in the drain current or the threshold voltage. The step heights in combination with the emission times were found to be characteristic for each defect. As a consequence, they can be used to identify each emission event with a particular defect. Based on these data, TDDS has provided a considerable amount of information on the defect behavior, suggesting an improved physical picture of the charge trapping phenomenon. The relevant observations made using TDDS are listed below:

- Charge capture and emission obey the statistics of first-order processes and were tentatively associated with a nonradiative multiphonon (NMP) process [61], as proposed by Kirton [10].

- Both charge capture and emission typically show a pronounced temperature dependence illustrated in Fig. 1. The extracted activation energies lie between 0.4 eV and 1.5 eV for the experimental time window used. It is noted that this range might substantially extend for increased time windows.

The pronounced temperature behavior suggests a strongly thermally activated C/E process, which is characteristic for an NMP process. However, this behavior is incompatible with previously proposed models, such as the elastic electron tunneling or the Shockley-Read-Hall model [63].

- Charge capture exhibits an exponential gate bias dependence over a wide voltage range, a fact that is also consistent with NMP processes but inconsistent with the SRH model (see Fig. 1).
- Charge capture is found to slow down for high-frequency AC gate bias [64]. Such a behavior indicates the existence of additional metastable states, which are involved in the charge capture process.
- The two defect variants which have so far been proposed differ significantly in their bias dependence [65]. The charge emission times of the "fixed oxide traps" are insensitive to the gate bias, even deep into accumulation (cf. Fig. 1). By contrast, "switching oxide traps" have emission times which vary at small gate biases. This again demonstrates that the observed C/E dynamics cannot rely on a simple two-state model and must involve metastable states.
- TDDS studies have revealed that electron capture and emission exhibit similar gate bias and temperature dependences as the fixed and switching hole traps in pMOSFETs [66]. Interestingly, the same trapping behavior was also observed in device technologies based on high-κ dielectrics [67, 68]. This suggests that charge trapping in those cases rests upon the same general concepts with similar physical processes being involved.

All the above findings led to the development of the four-state NMP model, which relies on the idea of a bistable defect in both charge states and is illustrated in the state diagram of Fig. 2. It was initially motivated by the Harry-Diamond-Laboratory (HDL) [69] model, whose three states were extended by a fourth one in order to explain the full richness of experimental features. Interestingly, these states would also be consistent with the configurations for the E' center proposed in ref. [70].

In the four-state NMP model, the defect can become charged or discharged during a capture or emission event and must, therefore, be described by two stable charge states (1,2) — such as in a simple two-state defect model. The peculiarity of this model, however, is the additional metastable states, which are present in both charge states and marked by primes ($1'$,$2'$). These metastable states are assumed to differ significantly in their configuration from their stable counterparts in the same charge state. As such, the transitions $1 \leftrightarrow 1'$ and $2 \leftrightarrow 2'$ are thought to be accompanied by large structural rearrangements, leading to defect deformation and sometimes proceed over barriers that can only be overcome by thermal activation.

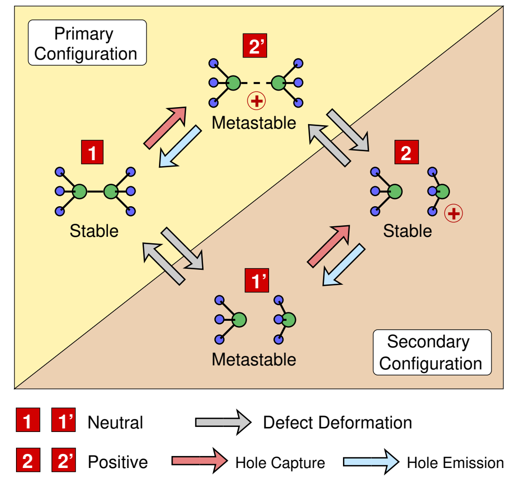


FIG. 2. State diagram of the multi-state model for an E' center. The defect has a stable neutral (1) and a stable positive (2) charge state, where each may have a metastable state marked by a prime ($1'$,$2'$). The NMP transitions $1 \leftrightarrow 2'$ and $1' \leftrightarrow 2$ occur between different charge states while the thermal transitions $1 \leftrightarrow 1'$ and $2 \leftrightarrow 2'$ proceed/take place between same charge states. It is noted that the transitions between the stable states are of primary interest as they correspond to the measured capture and emission times seen in TDDS experiments. However, they involve metastable states ($1'$,$2'$), which strongly affect the gate-bias and temperature dependence of the overall transition. The stick-and-ball models correspond to the atomic configurations of a possible defect candidate, shown here for illustration purposes only.

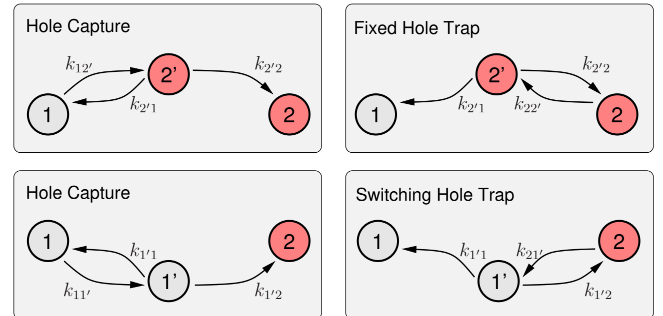


FIG. 3. Simplified state diagrams for hole capture (**top left**, **bottom left**) and for hole emission of a fixed (**top right**) and a switching (**bottom right**) hole trap. While hole capture primarily proceeds over state $2'$, hole emission is dominated by a transition over either of the states $1'$ or $2'$. For the latter, the time-limiting step is assumed to be the pure thermal transition between the states 2 and $2'$, which involves no CT and is gate bias independent. The former, however, is dominated by the CT reaction between the states $1'$ and 2, resulting in a strong gate bias dependence.

By contrast, the actual charge capture and emission process occurs between different charge states but involves a small structural rearrangement compared to that during the transitions $1 \leftrightarrow 1'$ and $2 \leftrightarrow 2'$.

Both, the defect deformation together with the NMP transition, constitutes the core of the four-state NMP

model illustrated in Fig. 2. The whole charge capture or emission consists of a two-step process, where the defect does not undergo a direct transition between the stable states (1,2) but takes the pathway over one of the metastable states ($1'$,$2'$) as indicated in Fig. 3. The state pairs (1, $2'$) and ($1'$,2) are grouped together under the term "primary" and the "secondary" configuration, respectively. It is emphasized here that this particular defect property ultimately allows for the trapping dynamics seen for the fixed and the switching oxide defects.

Even though these first TDDS studies have already provided deep physical insight into the mechanisms behind charge trapping, later long-time TDDS studies have augmented the picture of the bistable defect. The new findings are briefly summarized below [71, 72]:

- Defects occasionally disappear, remain inactive, and reappear after a certain amount of time. This phenomenon is denoted "defect volatility" and observed for the neutral and the positive charge state of the defects. Under switched bias conditions, this occurs for a large fraction of the defects and on widely distributed time scales, ranging from hours up to weeks. Importantly, there are no indications that the volatile defects behave otherwise differently than their non-volatile counterparts.
- Within the experimental error, the extracted deactivation and re-activation times are likely to follow an exponential distribution, consistent with a reaction-limited process. The volatility appears to be linked to a hydrogen reaction, as suggested by measurements on large-area devices with a different hydrogen content in the dielectric [73]. Hence, the sought defect was suggested to be a hydrogen-complexed defect.

These findings suggest a refined picture, in which the four-state NMP model is extended by additional inactive states (see the state diagram in Fig. 4). These newly introduced states exist for the neutral and the positive charge state and can be accessed via a reaction with hydrogen. Even though the reactants and their detailed reaction kinetics are still under investigation, the hydrogen bridge and the hydroxyl E' center appear to be promising candidates [74]. The latter is used to explain the state diagram of the extended NMP model in Fig. 4. In its neutral charge state (1), the hydroxyl E' consists of a hydroxyl (OH) group facing a Si dangling bond. This defect may be de-activated via a reaction to one of its precursor configurations: One possibility is that a free hydrogen atom saturates the dangling bond, resulting in the configuration 0^2. Alternatively, the hydrogen atom of the hydroxyl group can be released, leading to a distorted O-Si-O bridge in the pristine SiO_2 network. This bridge site is a precursor, called 0, since it is likely to form a hydroxyl E' center if a hydrogen atom passes by. Furthermore, the hydroxyl E' center could also be deactivated in its positive charge state $2'$. There, the hydrogen atom may attach to one of the nearby oxygen atoms, where it forms a fixed positive defect complex. Although further transitions to other inactive defect configurations are still conceivable, the concept of the extended NMP model relies on these inactive states which do not allow for hole capture or emission

The mentioned TDDS studies suggested a microscopic model for a hypothetical defect, whose atomic structure has remained unspecified. In order to confirm this model theoretically, the capture and emission times must be calculated from the energetics of a hypothetical defect and then compared to the time constants extracted from TDDS experiments. For this purpose, a detailed mathematical model is needed, which will be presented in the following section.

III. DERIVATION OF THE DEFECT MODEL

As charge capture and emission processes are at the heart of the four-state NMP model, calculating rates of these processes is vital for identifying defects responsible for these processes. For this purpose, the theory of CT reactions [41, 75, 76] provides an excellent basis. Its derivation will, therefore, be summarized in this section. Several essential aspects, such as the distinction between adiabatic and diabatic transitions, the Franck-Condon principle, the estimation of the electronic matrix elements, and the computation of the Franck-Condon factors will be discussed from the perspective of oxide traps in MOS transistors. The resulting CT rates will be incorporated within the framework of the four-state NMP model, which also includes thermally activated defect transformations. Furthermore, the CT rates will be generalized to account for the fact that the defect exchanges charge carriers with a full spectrum of electronic states from, for instance, the conduction or the valence band. Finally, Markov theory will be employed to derive the capture and emission times observed in experiments.

III.1. Adiabatic Charge Transfer Reactions

The most general formulation of CT reactions relies on a quantum mechanical description of the atomic system in which the reaction is taking place. Hence, the derivation of their corresponding rates starts from the time-independent molecular Schrödinger equation

$$\mathcal{H}_{\mathrm{mol}}|\Psi(\mathbf{r},\mathbf{R})\rangle = U|\Psi(\mathbf{r},\mathbf{R})\rangle \ . \tag{3.1}$$

Here, $\Psi(\mathbf{r},\mathbf{R})$ denotes the wavefunction of this system. The degrees of freedoms are given by the $3M$ electronic and $3N$ nuclei coordinates with M and N being the number of consider electrons and nuclei, respectively. U is the total energy of the atomic system, which is described by the following molecular Hamiltonian $\mathcal{H}_{\mathrm{mol}}$

$$\begin{aligned}\mathcal{H}_{\mathrm{mol}} = {} & \mathcal{T}_{\mathrm{n}}(\mathbf{R}) + \mathcal{T}_{\mathrm{e}}(\mathbf{r}) \\ & + \mathcal{V}_{\mathrm{ee}}(\mathbf{r}) + \mathcal{V}_{\mathrm{en}}(\mathbf{r},\mathbf{R}) + \mathcal{V}_{\mathrm{nn}}(\mathbf{R}) \ .\end{aligned} \tag{3.2}$$

It includes the standard terms for the electronic ($\mathcal{T}_{\mathrm{e}}(\mathbf{r})$) and the nuclear ($\mathcal{T}_{\mathrm{n}}(\mathbf{R})$) kinetic energies as well as

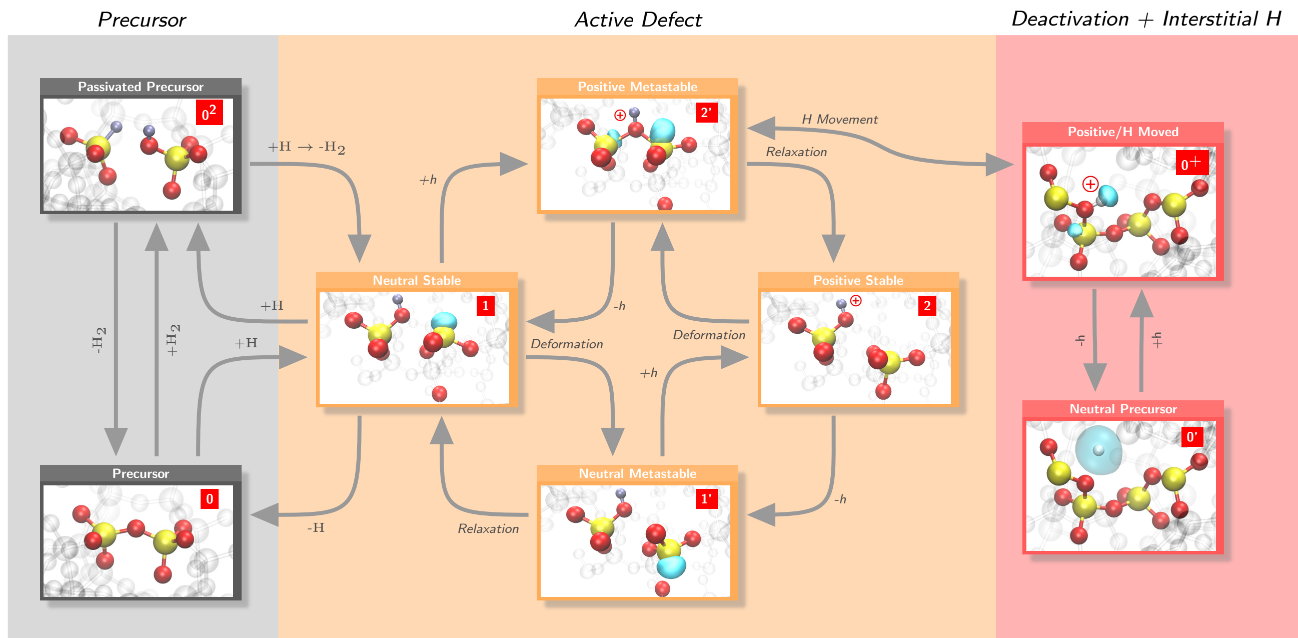


FIG. 4. Extended four-state NMP model illustrated for a promising defect candidate, the hydroxyl E' center. Still, the core of this model (middle) is build around the bistable defect with four states (1,1',2,2') and describes the active defect, which is capable of capturing and emitting charge carriers. However, the extended variant of the model also accounts for the inactive phases of the defect via transitions to the precursor states 0 and 0^2 (left) and the inactive states 0^+ and 0^0 (right).

the Coulombic contributions from the electron-electron ($\mathcal{V}_{\mathrm{ee}}(\mathbf{r})$), the electron-nuclei ($\mathcal{V}_{\mathrm{en}}(\mathbf{r},\mathbf{R})$), and the nuclei-nuclei ($\mathcal{V}_{\mathrm{nn}}(\mathbf{R})$) interactions.

In order to describe a CT reaction, the atomic system of the molecular Schrödinger equation (3.1) has to contain the atomic structure of the donor, which emits the transferred electron, as well as that of the acceptor, which captures the transferred electron. The resulting Schrödinger equation allows for the calculation of the quantum mechanical states which are required to evaluate the charge capture and emission rates. Considering charge trapping in transistors, this atomic system would consist of the substrate and the dielectric including the defect that captures or emits the charge carrier. If hole capture is considered, the donor corresponds to the defect and the acceptor to the substrate. Naturally, the roles of the donor and the acceptor are reversed for hole emission. For such large structures, the resulting Schrödinger equation represents a many-body problem of interacting electrons and nuclei. In general, this problem is mathematically intractable due to the complex correlations between the interacting particles but is even further complicated in this case due to the high-dimensionality of the considered atomic system and the unknown amorphous phase of the oxide material. However, the problem can be treated by two different approaches which make an assumption on the corresponding wavefunctions and are suited for different physical situations.

The common approach to solve the molecular Schrödinger equation is to use the adiabatic approximation, which is also referred to as the Born-Oppenheimer approximation. It is based on the assumption that electrons move on much faster than nuclei and can instantaneously adapt to each nuclei configuration. Therefore the nuclei coordinates can be treated as parameters in the total wavefunction and the adiabatic wavefunction can be approximately expressed as a product of an electronic ($\phi_i^{\mathrm{a}}(\mathbf{r};\mathbf{R})$) and a nuclear ($\eta_{i\alpha}^{\mathrm{a}}(\mathbf{R})$) wavefunctions:

$$\Psi_{i\alpha}^{\mathrm{a}}(\mathbf{r},\mathbf{R}) = \phi_i^{\mathrm{a}}(\mathbf{r};\mathbf{R})\eta_{i\alpha}^{\mathrm{a}}(\mathbf{R}) \; . \tag{3.3}$$

Employing this ansatz allows one to split the Schrödinger equation into a set of two coupled equations. The electronic Schrödinger equation is:

$$\mathcal{H}_{\mathrm{e}}|\phi_i^{\mathrm{a}}(\mathbf{r};\mathbf{R})\rangle = V_i^{\mathrm{a}}(\mathbf{R})|\phi_i^{\mathrm{a}}(\mathbf{r};\mathbf{R})\rangle \; , \tag{3.4}$$

with the Hamiltonian ($\mathcal{H}_{\mathrm{e}}$) defined by

$$\mathcal{H}_{\mathrm{e}} = \mathcal{T}_{\mathrm{e}}(\mathbf{r}) + \mathcal{V}_{\mathrm{ee}}(\mathbf{r}) + \mathcal{V}_{\mathrm{en}}(\mathbf{r};\mathbf{R}) + \mathcal{V}_{\mathrm{nn}}(\mathbf{R}) \; . \tag{3.5}$$

In the basis of the adiabatic wavefunctions, the Hamiltonian can be rewritten as

$$\mathcal{H}_{\mathrm{e}} = \sum_i V_i^{\mathrm{a}}(\mathbf{R}) \; |\phi_i^{\mathrm{a}}(\mathbf{r};\mathbf{R})\rangle\langle\phi_i^{\mathrm{a}}(\mathbf{r};\mathbf{R})| \; . \tag{3.6}$$

Since it is of a pure diagonal form, it does not allow for transitions between single electronic states i according to perturbation theory.

We note that equation (3.4) depends parametrically on the nuclei coordinates $\mathbf{R}$, which is why the solutions $V_i^{\mathrm{a}}(\mathbf{R})$ of the electronic Schrödinger equation are also functions of the nuclei coordinates $\mathbf{R}$. These solutions $V_i^{\mathrm{a}}(\mathbf{R})$ act as potentials for the nuclear motion in the coupled equation for nuclei:

$$\big(\mathcal{T}_{\mathrm{n}}(\mathbf{R}) + V_i^{\mathrm{a}}(\mathbf{R})\big)|\Psi_{i\alpha}^{\mathrm{a}}(\mathbf{r},\mathbf{R})\rangle = U_{i\alpha}^{\mathrm{a}}|\Psi_{i\alpha}^{\mathrm{a}}(\mathbf{r},\mathbf{R})\rangle \tag{3.7}$$

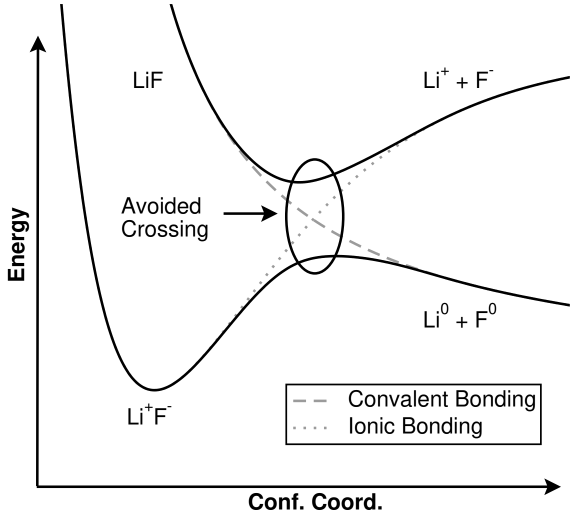


FIG. 5. Schematic configuration coordinate diagram of an adiabatic transition, illustrated for the system of a Li and a F atom. The solid black lines depict the adiabatic potential of the ground (lower curve) and the first excited (upper curve) state while the dashed lines show the behavior of the LiF system which preserves its bonding character, i.e. the ionic or covalent bonding. The former are specific to the Born-Oppenheimer approximation, which leads to an avoided crossing between the involved states and thus results in a charge transfer.

and are referred to as the adiabatic potential energy surface, discussed later in this section. For an N atoms system, this potential is a complex $3N$-dimensional hypersurface in the space of the nuclei coordinates $\mathbf{R}$. For illustration purposes, however, it is usually depicted along a particular configuration coordinate in 2D diagrams.

When the above molecular Hamiltonian is represented in the basis of the adiabatic wavefunctions, it reads

$$\begin{aligned}\mathcal{H}_{\mathrm{mol}} = &\sum_{i\alpha} U^{\mathrm{a}}_{i\alpha}\ |\Psi^{\mathrm{a}}_{i\alpha}(\mathbf{r},\mathbf{R})\rangle\langle\Psi^{\mathrm{a}}_{i\alpha}(\mathbf{r},\mathbf{R})| \\ &+ \sum_{i\alpha,j\beta} \theta^{\mathrm{a}}_{ij}(\mathbf{R})\ |\Psi^{\mathrm{a}}_{i\alpha}(\mathbf{r},\mathbf{R})\rangle\langle\Psi^{\mathrm{a}}_{j\beta}(\mathbf{r},\mathbf{R})|\ . \end{aligned} \quad (3.8)$$

This matrix is dominated by its diagonal elements $U^{\mathrm{a}}_{i\alpha}$ while the off-diagonal elements $\theta^{\mathrm{a}}_{ij}(\mathbf{R})$ are usually small except for rare cases discussed below.

The dominant elements $U^{\mathrm{a}}_{i\alpha}$ provide a good approximative solution for the stationary state of the molecular Hamiltonian. For instance, this is the case for the frequently discussed example of LiF bonding [77] shown schematically in Fig. 5. The LiF molecule is (highly) polar near the equilibrium distance. However, when the atoms separate to about 1 nm, the system of the Li and F atoms becomes much lower in energy than that of the Li^+ and F^- ions (see Fig. 5). During such a dissociation process, an electron is transferred from the F^- ion (donor) to the Li^+ ion (acceptor). As shown in the configuration coordinate diagram of Fig. 5, in the adiabatic approximation the molecule remains on the lower adiabatic potential instead of keeping its ionic bonding character (the dashed line). The transition rates along the adiabatic potential can be calculated using transition state theory and assume the form of an Arrhenius-type expression [40, 78, 79]

$$k^{\mathrm{a}} = \nu_0\ \exp(-\beta E^{\ddagger})\ . \quad (3.9)$$

Here, ν_0 takes the role of an attempt frequency and β is defined as $1/(k_{\mathrm{B}}T)$ with k_{B} and T being the Boltzmann constant and the temperature, respectively. $E^{\ddagger}$ corresponds to a thermal barrier, which must be surmounted for a transition to occur. Such transitions are referred to as “adiabatic” transitions and take place in many molecular systems where the donor and the acceptor are not too widely separated [80–82].

Similar description is applicable to defects in solids. The recent theoretical investigation [83] demonstrates that the donor-acceptor distance between defects may even extend up to the nanometer regime. Assuming that localized defects can be described by molecular models, it is conceivable that adiabatic transitions may also occur for defects located close to the substrate-dielectric interface of a transistor. Intriguingly, the adiabatic transitions exhibit a pronounced temperature and gate-bias dependence, as seen in the experimental RTN and TDDS data. In the four-state NMP model, the thermal behavior is due to the exponential temperature dependence in the Arrhenius-type equation (3.9) and is governed by the thermal barrier $E^{\ddagger}$. The gate-bias dependence originates from the fact that a CT reaction inevitably involves a dipole moment, such as the $\mathrm{Li}^+\mathrm{F}^-$ complex of Fig. 5. This dipole moment varies with the applied electric field and results in a field dependence of the adiabatic potential and the corresponding thermal barrier $E^{\ddagger}$ [6].

The Hamiltonian (3.8) has also additional matrix elements termed dynamical coupling elements $\theta^{\mathrm{a}}_{ij}(\mathbf{R})$ [84] in equation (3.8) that can become dominant for C/E processes. They originate from the fact that the spatial derivatives of the quantum mechanical nuclei momenta act on the electronic wavefunctions. In general, these elements have two contributions

$$\theta^{\mathrm{a}}_{ij}(\mathbf{R}) = \theta^{\mathrm{a}1}_{ij}(\mathbf{R}) + \theta^{\mathrm{a}2}_{ij}(\mathbf{R})\ . \quad (3.10)$$

The first term $\theta^{\mathrm{a}1}_{ij}(\mathbf{R})$ results from the second derivatives of the electronic wavefunctions $\phi^{\mathrm{a}}_i(\mathbf{r};\mathbf{R})$ with respect to the nuclei coordinates $\mathbf{R}$ and usually yields a negligible correction to the adiabatic potential and will not be discussed in further detail here. The second term reads [41, 77]

$$\theta^{\mathrm{a}2}_{ij}(\mathbf{R}) = -\sum_n \frac{\hbar^2}{M_n}\langle\eta^{\mathrm{a}}_{i\alpha}(\mathbf{R})|\mathbf{d}(\mathbf{R})|\nabla_n\eta^{\mathrm{a}}_{j,\beta}(\mathbf{R})\rangle \quad (3.11)$$

$$\mathbf{d}(\mathbf{R}) = \langle\phi^{\mathrm{a}}_i(\mathbf{r};\mathbf{R})|\nabla_n\phi^{\mathrm{a}}_j(\mathbf{r};\mathbf{R})\rangle\ . \quad (3.12)$$

Here, n denotes the nuclei coordinate and M_n corresponds to the respective nuclei mass and ∇_n to the respective spatial derivative. This term plays an important role as it couples the nuclei states α and β via nonadiabatic transitions. The farther the donor and the acceptor

are separated, the smaller the splitting becomes around the avoided crossing. As a consequence, the dynamical coupling element $\theta_{ij}^{\mathrm{a}2}(\mathbf{R})$ begins to diverge, leading to the breakdown of the Born-Oppenheimer approximation. This implies that the donor complex in the state α is likely to transfer to the different state β in the region around the avoided crossing. In the context of dissociation of the LiF molecule, such transitions tend to preserve the bonding character and are represented by the dashed $\mathrm{Li}^+ + \mathrm{F}^-$ potential in Fig. 5.

III.2. Nonadiabatic Charge Transfer Reactions

For an atomic system where the donor and acceptor are far separated, CT reactions are better treated within the diabatic approximation. In [40, 41] this approximation was employed for heterogeneous CT reactions, which can take place at the interface between a solid (a semiconductor or a metal [85]) and an electrolyte [40], consisting of molecules. From a quantum mechanical point of view, one faces a transition from a localized to a delocalized state or vice versa. An analogous situation is found for charge capture and emission in transistors where the defect and bulk wavefunctions play the roles of the localized and the delocalized states, respectively. As such, the diabatic approximation appears to be the more suitable approach and will, therefore, be used to derive the so-called 'nonadiabatic' CT reactions [41, 76, 78].

The generalized formulation of the nonadiabatic CT reactions rests upon an adiabatic-to-diabatic transformation [86]

$$\phi_i^{\mathrm{a}}(\mathbf{r},\mathbf{R}) = S_{ij}(\mathbf{R})\phi_j^{\mathrm{d}}(\mathbf{r},\mathbf{R}) \ , \tag{3.13}$$

$S_{ij}(\mathbf{R})$ denotes a unitary operator and the superscripts 'a' and 'd' indicate quantities defined within the adiabatic or the diabatic approximation, respectively. This transformation allows to construct a basis set which minimizes the singularities of the dynamical coupling elements and thus removes the avoided crossing [87, 88].

$$\theta_{ij}^{\mathrm{a}2}(\mathbf{R}) \approx 0 \tag{3.14}$$

However, the resulting set of diabatic wavefunctions $\phi_i^{\mathrm{d}}(\mathbf{r};\mathbf{R})$ does not diagonalize the electronic Hamiltonian

$$\begin{aligned}\mathcal{H}_{\mathrm{e}} = &\sum_i V_i^{\mathrm{d}}(\mathbf{R})\ |\phi_i^{\mathrm{d}}(\mathbf{r};\mathbf{R})\rangle\langle\phi_i^{\mathrm{d}}(\mathbf{r};\mathbf{R})| \\ &+ \sum_{i\neq j} \theta_{ij}^{\mathrm{d}}(\mathbf{R})\ |\phi_i^{\mathrm{d}}(\mathbf{r};\mathbf{R})\rangle\langle\phi_j^{\mathrm{d}}(\mathbf{r};\mathbf{R})| \ .\end{aligned} \tag{3.15}$$

While this Hamiltonian is still dominated by its diagonal elements $V_i^{\mathrm{d}}(\mathbf{R})$, new off-diagonal elements $\theta_{ij}^{\mathrm{d}}(\mathbf{R})$ appear within the diabatic approximation. The diagonal elements correspond to the diabatic potential (see Fig. 6) of the electronic state i and usually match well with their adiabatic counterparts — except for the regions around avoided crossings. There, the diabatic potentials intersect and are coupled via the small off-diagonal elements, defined by the static coupling element [89]

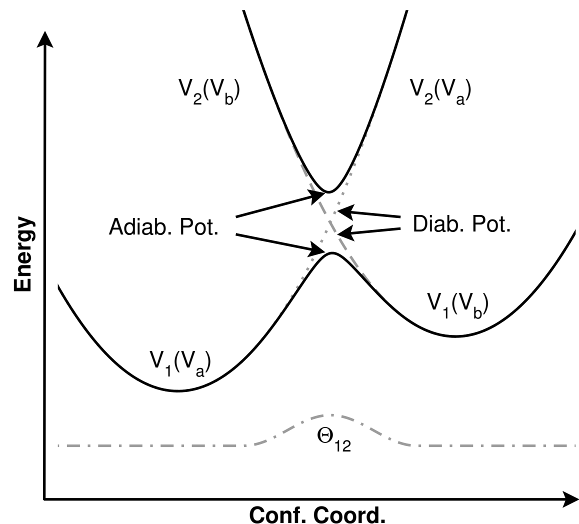


FIG. 6. Comparison between the adiabatic (solid line) and the diabatic (dashed line) representation of a CT reaction. Their corresponding electronic states are labeled by numbers (1,2) or letters (a,b), respectively. A large dynamical coupling element Θ_{12}^{d} leads to an avoided crossing within the adiabatic approximation, where the CT reaction proceeds over the barrier of the adiabatic potential V_1. By contrast, a large static coupling element Θ_{12}^{a} allows for transitions between the potentials V_1 and V_2 in the vicinity of the avoided crossing. This case is then better described within the diabatic approximation with the intersecting diabatic potentials V_a and V_b.

$$\theta_{ij}^{\mathrm{d}}(\mathbf{R}) = \langle\phi_i^{\mathrm{d}}(\mathbf{r};\mathbf{R})|\mathcal{H}_{\mathrm{e}}|\phi_j^{\mathrm{d}}(\mathbf{r};\mathbf{R})\rangle \ . \tag{3.16}$$

As the construction of a diabatic basis set (3.13) is not well defined by equation (3.14), there exist no accurate analytical expressions for this quantity. However, a simplifying tight-binding approach leads to coupling elements which show a strong exponential decay with an increasing separation between the acceptor and the donor [41]. This exponential behavior has been confirmed by highly sophisticated simulations based on constrained DFT (CDFT) [80, 81, 90, 91] and also holds for defects in solids according to more recent CDFT studies [82, 83]. Following the argumentation in [41, 82], this behavior can be traced back to the exponential dependence of electron tunneling and in further consequence to the exponential decay of the electron wavefunctions involved.

The static coupling terms also appear in the molecular Hamiltonian

$$\begin{aligned}\mathcal{H}_{\mathrm{mol}} = &\sum_{i\alpha} U_{i\alpha}^{\mathrm{d}}\ |\Psi_{i\alpha}^{\mathrm{d}}(\mathbf{r},\mathbf{R})\rangle\langle\Psi_{i\alpha}^{\mathrm{d}}(\mathbf{r},\mathbf{R})| \\ &+ \sum_{i\alpha\neq j\beta} \theta_{ij}^{\mathrm{d}}(\mathbf{R})\ |\Psi_{i\alpha}^{\mathrm{d}}(\mathbf{r},\mathbf{R})\rangle\langle\Psi_{j\beta}^{\mathrm{d}}(\mathbf{r},\mathbf{R})|\end{aligned} \tag{3.17}$$

and lead to a coupling in the molecular Hamiltonian, associated with 'nonadiabatic' CT reactions. In comparison to equation (3.8), the dynamic coupling terms $\theta_{ij}^{\mathrm{a}}(\mathbf{R})$ are absent in the diabatic representation now. This results from the fact that the diabatic wavefunctions are supposed to vary weakly with the nuclei coordinates due to the condition imposed by equation (3.14).

III.3. Fermi's Golden Rule

As pointed out in the previous section, the diabatic representation is more suited for physical situations where the static coupling between the electronic states is small and the transition occurs at the intersection of the diabatic potentials. For instance, this case is encountered for molecular junctions [40], where a chain of consecutive CT reactions through one or more molecules between two electrodes yields a current. These CT reactions involve the delocalized bulk states of the substrate or the metal gate and a localized state of a molecule. Both states are far separated at atomic scales, resulting in a large tunneling barrier. Therefore, the static coupling elements $\theta^{\mathrm{d}}_{ij}(\mathbf{R})$ become small and Fermi's golden rule can be employed to calculate the nonadiabatic CT rates $k_{i\alpha,j\beta}$

$$k_{i\alpha,j\beta} = \frac{2\pi}{\hbar}|M_{i\alpha,j\beta}|^2\delta(U_{i\alpha} - U_{j\beta}) \tag{3.18}$$

$$M_{i\alpha,j\beta} = \langle\eta^{\mathrm{d}}_{i\alpha}(\mathbf{R})|\theta^{\mathrm{d}}_{ij}(\mathbf{R})|\eta^{\mathrm{d}}_{j\beta}(\mathbf{R})\rangle\ . \tag{3.19}$$

The expression $\delta(U_{i\alpha} - U_{j\beta})$ in equation (3.18) corresponds to the Dirac-delta function and imposes energy conservation ($U_{i\alpha} \approx U_{j\beta}$) during the CT reaction. The electronic matrix element $\theta^{\mathrm{d}}_{ij}(\mathbf{R})$ takes a central role in the theory of CT reactions since it determines the magnitude of the transition rate $k_{i\alpha,j\beta}$.

The CT reactions involved in the conduction of molecular junctions have a lot of similarities with the C/E processes seen in microelectronic devices. In both cases, there is a wavefunction of a delocalized bulk state, which decays exponentially with an increasing distance from the interface, and a localized wavefunction, which belongs to either a molecule or a defect, respectively. Furthermore, the bulk and defect are also widely separated on atomic scales, analogously to the electrodes and the molecules at the molecular junctions. Therefore the theoretical framework used for describing the nonadiabatic CT reactions in molecular junctions provides a basis for calculating rates for the C/E processes in the investigated charge trapping phenomena.

For the sake of completeness, we note that there are also alternative methods to determine CT rates. One approach relies on the static approximation, where the electronic wavefunctions are evaluated for fixed nuclear coordinates. Since these wavefunctions do not depend on the atomic configuration, the static approximation [92] can be regarded as a special case of the diabatic approximation. Recently, Alkauskas *et al.* [93] have demonstrated that this method yields surprisingly accurate results for bulk defects in semiconductors using conventional DFT calculations. Another approach is based on a semi-classical picture [94] where the electronic matrix element takes the form of the exponential term in equation (3.9), but now with the energy barrier measured up to the intersection point. This semi-classical theory can be improved by treating the electronic transition within the Landau-Zener theory. Even though the resulting expression covers the diabatic as well as the adiabatic regime, it still neglects the effect of nuclear tunneling. For the remainder of this paper, we use the rate expression (3.18) since it is the most general formulation of nonadiabatic transitions.

III.4. Franck-Condon Principle

For the determination of CT rates (3.18), the largest effort is required to calculate the matrix elements $M_{i\alpha,j\beta}$. Since the static coupling element $\theta^{\mathrm{d}}_{ij}(\mathbf{R})$ is a function of the nuclei coordinates $\mathbf{R}$, it must be evaluated for each configuration. In principle, this task can be achieved using CDFT. For C/E processes, however, the acceptor/-donor complex consists of a semiconductor or metal on the one hand and an insulator containing the defect on the other hand. Therefore, a realistic atomistic model must include at least several hundreds of atoms, resulting in computationally unfeasible costs for CDFT. Alternatively, the static approximation allows one to rigorously treat the dependence on the nuclei coordinates by use of the q-centroid approximation [75] but also this method exceeds the computational capabilities for the large atomic systems of interest. Therefore, analytical expressions are usually applied for the evaluation of the electronic matrix element [95].

As a reasonable approximation, the electronic wavefunctions can be assumed to vary weakly with the nuclei coordinates. For capture and emission in charge trapping, this statement can be justified by the fact that the defect distortions do not significantly affect the defect wavefunction in the region of its exponential decay. Based on this assumption, the matrix element $M_{i\alpha,j\beta}$ can be simplified to

$$M_{i\alpha,j\beta} = \theta^{\mathrm{d}}_{ij}(\mathbf{R}) I_{i\alpha,j\beta}\ , \tag{3.20}$$

where

$$I_{i\alpha,j\beta} = \langle\eta^{\mathrm{d}}_{i\alpha}(\mathbf{R})|\eta^{\mathrm{d}}_{j\beta}(\mathbf{R})\rangle \tag{3.21}$$

are the so-called Franck-Condon factors. The above approximation is widely used in the field of spectroscopy and quantum chemistry [96] and is usually referred to as the Franck-Condon principle [96, 97].

The Franck-Condon factor $I_{i\alpha,j\beta}$ is given by the overlap integral of the nuclear wavefunctions of the initial ($i\alpha$) and the final ($j\beta$) states of the system in equation (3.20). Both are obtained from their respective molecular Schrödinger equations:

$$\mathcal{H}_{\mathrm{mol},k} = \sum_n \frac{\mathcal{P}_n^2}{2M_n} + V_k(\mathbf{R}) \tag{3.22}$$

$$\mathcal{H}_{\mathrm{mol},k}\, \eta_{k\gamma}(\mathbf{R}) = U_{k\gamma}\, \eta_{k\gamma}(\mathbf{R})\ , \tag{3.23}$$

where $\mathcal{P}_n$ denotes the quantum mechanical momentum operator of the nuclei n. The electronic states i and j are subsumed by the index k and the nuclei state α and β by the index γ. $V_k(\mathbf{R})$ still corresponds to the

diabatic potential but the superscript 'd' will be omitted for the remainder of the paper. It governs the nuclear motion, which usually has an oscillatory behavior in solids. These oscillations corresponds to lattice vibrations and close to minima can be described within the harmonic approximation. Accordingly, the diabatic potential $V_k(\mathbf{R})$ can be expanded into a Taylor series around the energy minimum configuration $R^0_{k,n/m}$ and truncated after the second-order terms. Then the molecular Schrödinger equation reads

$$\mathcal{H}_{\mathrm{mol},k} = \sum_{n,m} \frac{\mathcal{P}^2_{k,n}}{2M_n} + V^0_k + \frac{V^2_{k,nm}}{2} \Delta R_{k,n} \Delta R_{k,m} \tag{3.24}$$

$$\mathcal{H}_{\mathrm{mol},k}\, \eta_{k\gamma}(\mathbf{R}_k) = U_{k\gamma}\, \eta_{k\gamma}(\mathbf{R}_k) \ , \tag{3.25}$$

where $\Delta R_{k,n/m} = R_{k,n/m} - R^0_{k,n/m}$ are the components of the displacement vectors. The second and third terms in equation (3.24) describe the diabatic potential, which assumes a paraboloid shape. Its energy minimum is defined by

$$V^0_k = V_k(\mathbf{R}^0_k) \tag{3.26}$$

and its curvature is given by

$$V^2_{k,nm} = \left. \frac{\partial^2 V_k(\mathbf{R})}{\partial R_{k,n} \partial R_{k,m}} \right|_{\mathbf{R}^0_k} . \tag{3.27}$$

It is noted that the first-order terms in the Hamiltonian (3.24) vanish because no forces act on the nuclei at their energy minima per definition.

The diabatic potential $V^2_{k,nm}$ is of a nondiagonal form, leading to a coupling between the Hamiltonians of the nuclei coordinates n and m. In order to eliminate this coupling, the Schrödinger equation (3.24) is expressed in mass-weighted nuclear coordinates [41, 75]

$$\bar{Q}_{k,n} = \sqrt{M_n} R_{k,n} \tag{3.28}$$

$$\Delta\bar{Q}_{k,n} = \sqrt{M_n} \Delta R_{k,n} \tag{3.29}$$

and their corresponding nuclear momenta

$$\bar{\mathcal{P}}_{k,n} = \frac{\mathcal{P}_{k,n}}{\sqrt{M_n}} . \tag{3.30}$$

Using the above definitions, the molecular Schrödinger equation reads

$$\mathcal{H}_{\mathrm{mol},k} = \sum_{n} \frac{\bar{\mathcal{P}}^2_{k,n}}{2} + V^0_k + \sum_{n,m} \frac{V^2_{k,nm}}{2\sqrt{M_n M_m}} \Delta\bar{Q}_{k,n} \Delta\bar{Q}_{k,m} \tag{3.31}$$

$$\mathcal{H}_{\mathrm{mol},k}\, \eta_{k\gamma}(\Delta\bar{Q}_k) = U_{k\gamma}\, \eta_{k\gamma}(\Delta\bar{Q}_k) \ . \tag{3.32}$$

The tensor for the potential energy can be diagonalized via an orthogonal transformation while keeping the diagonal form of the kinetic energy tensor.

$$\mathcal{H}_{\mathrm{mol},s} = \frac{\bar{\mathcal{P}}^2_{k,s}}{2} + V^0_k + \frac{\omega^2_{ks}}{2} \Delta\bar{Q}^2_{k,s} \tag{3.33}$$

$$\Big\{ \sum_s \mathcal{H}_{\mathrm{mol},s} \Big\}\, \eta_{k\gamma s}(\Delta\bar{Q}_{k,s}) = U_{k\gamma s}\, \eta_{k\gamma s}(\Delta\bar{Q}_{k,s}) \tag{3.34}$$

Due to the diagonalization, the single Hamiltonians (3.33) are now decoupled and can therefore be treated separately. Each of them represents a quantum harmonic oscillator for the normal mode s with a vibrational frequency ω_{ks} defined by

$$\omega^2_{ks} = \frac{V^2_{k,nm}}{\sqrt{M_n M_m}} . \tag{3.35}$$

Six of the normal modes are zero-valued and correspond to rigid translation or rotation of the atomic structure. Since they do not affect the kinetics of the CT reaction, they will be omitted in all sums over the normal modes s in the rest of the paper. It is noted here that the nuclei coordinates can now be related to the vibrations of quantum harmonic oscillators and thus the term vibrational wavefunction will be preferred to nuclei wavefunction from now on.

The Hamiltonian (3.33) is convenient because solutions for the harmonic oscillator are well-known and therefore the above transformation is applied to the initial (i) as well as the final (j) molecular Schrödinger equation in order to obtain the vibrational wavefunctions $\eta_{i\alpha}$ and $\eta_{j\beta}$ required to determine the Franck-Condon factors in equation (3.20). These wavefunctions describe the defect vibrations before and after the CT reaction in terms of the initial and final normal modes.

We note, however, that these modes may actually differ. As a result, the decay of one mode can give rise to the excitation of several other modes during a CT reaction and thus leads to a mode mixing known as the Duschinsky effect [75, 98–101]. Since this effect only becomes relevant for a highly accurate determination of the CT rates, it is neglected in our studies. Despite this simplification, the expression of the CT rates still involves the overlap integrals of all combinations between the initial (s_i) and the final (s_j) normal modes. Since such an accurate approach would result in a high complexity of the calculations, the single-mode approximation is frequently proposed. It rests upon the assumption that the dominant normal modes for the initial and the final state coincide. This assumption has been implicitly employed in several works [102–104] but was recently justified by the work of Alkauskas *et al.* [105, 106]. Consequently, the molecular Schrödinger equations for the initial (i) and the final (j) state can be expressed by using only one nuclei coordinate Q, termed the configuration coordinate from

now on. They simplify to

$$\left\{-\frac{\hbar^2}{2}\frac{\partial^2}{\partial Q_i^2}+V_i^0+\frac{\omega_i^2}{2}\Delta Q_i^2\right\}\eta_{i\alpha}(\Delta Q_i) = U_{i\alpha}\eta_{i\alpha}(\Delta Q_i) \tag{3.36}$$

$$\left\{-\frac{\hbar^2}{2}\frac{\partial^2}{\partial Q_j^2}+V_j^0+\frac{\omega_j^2}{2}\Delta Q_j^2\right\}\eta_{j\beta}(\Delta Q_j) = U_{j\beta}\eta_{j\beta}(\Delta Q_j) \ , \tag{3.37}$$

with the well-known solutions

$$U_{i\alpha} = V_i^0 + (n_\alpha + 1/2)\,\hbar\omega_i \tag{3.38}$$

$$U_{j\beta} = V_j^0 + (n_\beta + 1/2)\,\hbar\omega_j \ . \tag{3.39}$$

The configuration coordinates ΔQ_i and ΔQ_j are referenced to their corresponding minimum energy configuration Q_i^0 and Q_j^0, respectively, which are displaced by ΔQ_{ij}

$$\Delta Q_i = Q - \underbrace{(-\Delta Q_{ij}/2)}_{Q_i^0} \tag{3.40}$$

$$\Delta Q_j = Q - \underbrace{(+\Delta Q_{ij}/2)}_{Q_j^0} \tag{3.41}$$

$$\Delta Q_{ij} = Q_j^0 - Q_i^0 = \Delta Q_i - \Delta Q_j \tag{3.42}$$

The above molecular Schrödinger equations represent two quantum mechanical harmonic oscillators, which have nonorthogonal vibrational wavefunctions due to their different vibrational frequencies and/or the displacement of their equilibrium positions. This fact gives rise to a wavefunction overlap and results in nonvanishing Franck-Condon factors in (3.20). These factors describe an effect similar to nuclear tunneling [78]. As shown in [107], the overlap of wavefunctions can be interpreted as a nuclear tunneling process. In this paper, the CT process was re-formulated in a theory which is based on nuclear tunneling and used as a description for the thermal ionization of deep impurities.

For the sake of completeness, we note that the curvatures ($\frac{1}{2}\omega_i$,$\frac{1}{2}\omega_j$) of the harmonic oscillators are usually specified by Huang-Rhys factors (S_i,S_j), defined by the equations:

$$\frac{\omega_i^2}{2}\Delta Q_{ij}^2 := S_i\hbar\omega_i \tag{3.43}$$

$$\frac{\omega_j^2}{2}\Delta Q_{ij}^2 := S_j\hbar\omega_j \ , \tag{3.44}$$

and are denoted as reorganization energies [41]. As illustrated in the configuration coordinate diagram of Fig. 7, the Huang-Rhys factor S_i corresponds to the energy difference between the energy points $V_i(Q_j^0)$ at P_{i2} and $V_j(Q_j^0)$ at P_{i1} and is expressed in multiples of the vibrational energy quanta $\hbar\omega_i$. Analogously, $S_j\hbar\omega$ is defined as the energy difference between P_{j2} and P_{j1}.

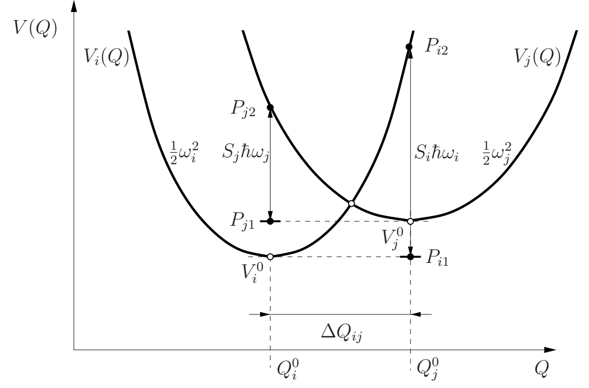


FIG. 7. Definition of the Huang-Rhys factors S_i and S_j. This figure shows the diabatic potentials for the initial (i) and the final (j) state with their shapes given by the harmonic potentials $V_i(\Delta Q_i) = V_i^0 + \omega_i^2\Delta Q_i^2/2$ and $V_j(\Delta Q_j) = V_j^0 + \omega_j^2\Delta Q_j^2/2$, respectively.

III.5. Thermally Activated Transitions

Even though the above description provides a well-suited method to determine the Franck-Condon overlap factors, energy conservation must still be met precisely due to the Dirac delta function in (3.18). This means that CT reactions only take place if the energies of the initial and the final states coincide. However, the CT reaction may also proceed over a thermally excited state $i\alpha$ of the molecular Hamiltonian in state i. According to the basic principles of statistical mechanics, the occupancy of an excited state is calculated by taking the thermal average

$$\underset{\alpha}{\text{ave}} = \sum_\alpha p_{i\alpha}(\Omega_{i\alpha}) \ , \tag{3.45}$$

where

$$p_{i\alpha} = \frac{\exp\left(-\frac{\Omega_{i\alpha}}{k_\mathrm{B}T}\right)}{\sum_\gamma \exp\left(-\frac{\Omega_{i\gamma}}{k_\mathrm{B}T}\right)} \tag{3.46}$$

is the statistical weight of the initial state $i\alpha$ of the CT reaction. All transitions into the final states j are considered by a summation over all final states β.

$$k_{i,j} = \frac{2\pi}{\hbar}|\theta_{ij}^\mathrm{d}(\mathbf{R})|^2 \underset{\alpha}{\text{ave}} \sum_\beta |I_{i\alpha,j\beta}|^2\delta(U_{i\alpha} - U_{j\beta}) \ . \tag{3.47}$$

Furthermore, the energies $U_{i\alpha}$ and $U_{j\beta}$ in the Dirac-delta function can be decomposed according to

$$U_{i\alpha} = V_i^0 + \Omega_{i\alpha} \tag{3.48}$$

$$U_{j\beta} = V_j^0 + \Omega_{j\beta} \ , \tag{3.49}$$

where $\Delta V_{ij} = V_j^0 - V_i^0$ denotes the energy shift between the diabatic potentials i and j and $\Omega_{i\alpha} = (n_i + 1/2)\hbar\omega_i$ and $\Omega_{j\beta} = (n_j + 1/2)\hbar\omega_j$ are the vibrational energies of

the harmonic oscillators. Using the above definitions, $k_{i,j}$ can be expressed as

$$k_{i,j} = \frac{2\pi}{\hbar}|\theta^{\mathrm{d}}_{ij}(\mathbf{R})|^2\xi_{ij} \tag{3.50}$$

with the quantity ξ_{ij} defined as

$$\xi_{ij} = \operatorname*{ave}_{\alpha}\sum_{\beta}|I_{i\alpha,j\beta}|^2\delta(\Omega_{i\alpha}-\Omega_{j\beta}-\Delta V_{ij}) \ . \tag{3.51}$$

According to this equation, the energy shift ΔV_{ij} must be compensated by the energy difference in the vibrational energies $\Omega_{i\alpha}-\Omega_{j\beta}$ in order for a CT reaction to be allowed.

III.6. Electronic Matrix Element

In the context of perturbation theory, the static coupling is frequently denoted as the electronic matrix element. Due to its importance for CT reactions, it has been the focus of numerous investigations [80, 82, 83, 90, 91]. From simple tight-binding derivations, the static coupling factor is expected to show an exponential distance dependence.

$$\theta^{\mathrm{d}}_{ij}(\mathbf{R}) = H_{ij}\exp(-\kappa d/2) \tag{3.52}$$

Here, H_{ij} corresponds to a simple prefactor, which can be estimated from CDFT simulations for molecular CT reactions [80, 90, 91]. κ is termed the decay constant and determines the exponential dependence of the static coupling factor on the donor-acceptor distance d. This exponential dependence has recently been confirmed for heterogeneous CT reactions, which also involve a bulk defect states [80, 82, 83].

The static coupling factor is also similar to the electronic matrix element in the nonradiative multi-phonon, NMP, theory, where different approximate expressions for tunneling factor were proposed [30, 36, 38]. They can be justified by the fact that the magnitude of the electronic matrix element is dominated by the exponential decay of the overlap between the defect and the bulk band wavefunction. Since the defect wavefunction is assumed to be strongly localized compared to the bulk wavefunction, the electronic matrix element reduces to a tunneling expression to first order. The electronic matrix element can then be estimated by

$$\theta^{\mathrm{d}}_{ij}(\mathbf{R}) \approx \langle\phi^{\mathrm{d}}_i(\mathbf{r},\mathbf{R})|\hat{H}_{\mathrm{e}}|\phi^{\mathrm{d}}_j(\mathbf{r},\mathbf{R})\rangle \tag{3.53}$$

$$\approx k\langle\phi^{\mathrm{d}}_i(\mathbf{r},\mathbf{R})|\phi^{\mathrm{d}}_j(\mathbf{r},\mathbf{R})\rangle \tag{3.54}$$

$$\approx k|\phi^{\mathrm{d}}_i(\mathbf{r}_{\mathrm{d}},\mathbf{R})|^2 \approx \tilde{k}\lambda(\mathbf{r}_{\mathrm{d}},\mathbf{R},E) \ , \tag{3.55}$$

where $\tilde{k}$ is a parameter that needs to be calibrated to experimental data and $\lambda(\mathbf{r}_{\mathrm{d}},E)$ represents the dimensionless Wenzel-Kramers-Brillouin, WKB, factor with E being the energy of the transfered charge carrier (see Appendix). With the above expression for the electronic matrix element, the CT transition rates can be rewritten as

$$k_{i,j} = \frac{2\pi}{\hbar}\tilde{k}^2\lambda^2(\mathbf{r}_{\mathrm{d}},E)\,\xi_{ij} \ . \tag{3.56}$$

III.7. Evaluation of the Franck-Condon Factors

The quantity ξ_{ij} in equation (3.51) corresponds to the spectral lineshape function [108], which has been studied in detail in the context of fluorescence and optical absorption [41, 96]. The corresponding optical spectrum features a comb of sharp peaks at sufficiently low temperatures. Its peculiar shape results from the discreteness of the energy levels involved in the optical transitions and has been derived for F centers in solids [96]. The underlying theory has also been extended for nonradiative transitions in solids [109], in which the characteristic comb is not observed for typical operation temperatures of transistors. This originates from the fact that the energies $U_{i\alpha}$ and $U_{j\beta}$ are broadened due to their finite lifetimes. The broadening can be described by replacing the Dirac-Delta function $\delta(U_{i\alpha}-U_{j\beta})$ in equation (3.47) with a more realistic Gaussian distribution

$$g(\Delta U) = \frac{1}{\sqrt{2\pi\sigma}}\exp\big(-\frac{\Delta U^2}{2\sigma^2}\big) \ . \tag{3.57}$$

Here, σ depends on the lifetime of the defect in the single states of the quantum mechanical harmonic oscillator. As these lifetimes are related to the complex interactions with the defect environment, this quantity is hard to determine and was estimated to be around the separation of the oscillator energies in this study.

Using this substitution, the lineshape function reads

$$\xi_{ij} = \operatorname*{ave}_{\alpha}\sum_{\beta}|I_{i\alpha,j\beta}|^2 g(\Omega_{i\alpha}-\Omega_{j\beta}-\Delta V_{ij}) \ , \tag{3.58}$$

where the magnitude of a certain peak is determined by the overlap $I_{i\alpha,j\beta}$ of the vibrational wavefunctions $i\alpha$ and $j\beta$. To the best of our knowledge, closed-form solutions of $I_{i\alpha,j\beta}$ only exist for the case that the initial and the final vibrational frequencies ($\omega_i=\omega_j$) are equal [40]. In the context of CT reactions, this assumption yields "linear electron-phonon coupling" and is frequently employed in the literature to compute the Franck-Condon factors because of its mathematical simplicity.

The more general case of differing vibrational frequencies is referred to as "quadratic electron-phonon coupling" but can only be solved numerically. One method is based on a finite-volume discretization for solving the molecular Schrödinger equation. Unfortunately, it suffers from the numerical inaccuracies in the exponential tails of the vibrational wavefunctions. Alternatively, the vibrational wavefunctions can be derived from the recurrence relations of the Hermite polynomials, however, this method only remains sufficiently accurate up to a few tens of iterations. Therefore, Schmidt *et al.* [110, 111]

have developed a new iteration scheme to directly evaluate the Franck-Condon overlap factors from an iteration scheme. This method was found to be accurate up to a few hundred iterations and has therefore been applied in all our studies. The expression (3.58) of the lineshape function simplifies significantly in the high-temperature limit, where $k_\mathrm{B}T$ exceeds the magnitude of the vibrational energy quantum $\hbar\omega_i$. In this regime, the energy separation of $\hbar\omega_i$ can be thermally overcome and the oscillator dynamics can be treated classically. This implies that the diabatic potentials $V_i(Q)$ and $V_j(Q)$ as functions of the configuration coordinate Q can be expressed as:

$$\begin{aligned} V_i(Q) &= c_i(Q - Q_i^0)^2 + V_i^0 \\ &= c_i {\Delta Q_i}^2 + V_i^0 \end{aligned} \tag{3.59}$$

$$\begin{aligned} V_j(Q) &= c_j(Q - Q_j^0)^2 + V_j^0 \\ &= c_j(\Delta Q_i - \Delta Q_{ij})^2 + V_i^0 + \Delta V_{ij} \end{aligned} \tag{3.60}$$

using the curvatures

$$c_i = \frac{1}{2}\omega_i^2 \tag{3.61}$$

and

$$c_j = \frac{1}{2}\omega_j^2 \tag{3.62}$$

of the diabatic potentials i and j, respectively. $V^0_{i/j}$ represents the energy minimum of the diabatic potential $V_{i/j}(Q)$ but corresponds to an additative constant that can be neglected without loss of generality. ΔV_{ij} is the energy difference between the initial and final states and determines the direction of the CT reaction. In the next chapter, this quantity will be related to trap energy levels, which are usually used to discuss C/E phenomena.

In the classical limit ($\hbar \to 0$), the discrete energy spectrum of $U_{i\alpha}$ and $U_{j\beta}$ becomes continous. Furthermore, the corresponding vibrational wavefunctions $\eta_{i\alpha}$ and $\eta_{j\beta}$ peak around their classical turning points Q^* defined by the condition $V_{i\alpha/j\beta}(Q^*) = U_{i\alpha/j\beta}$ and have only a significant overlap at the intersection point of the diabatic potentials $V_i(Q)$ and $V_j(Q)$. As a result, the Condon factors have to vanish in a classical equivalent of the lineshape function [112] and the allowed CT reactions are restricted to those transitions in which the diabatic potentials intersect. Then, the classical lineshape function can be simplified to:

$$\begin{aligned} &\xi_{ij}(c_i, c_j, \Delta Q_{ij}, \Delta V_{ij}) \\ &\quad = Z^{-1} \int\limits_Q \mathrm{e}^{-\beta U_i(Q')} \delta\left(V_i(Q') - V_j(Q')\right) \mathrm{d}Q' \end{aligned} \tag{3.63}$$

with the partition function

$$Z = \int\limits_Q \mathrm{e}^{-\beta U_i(Q')} \mathrm{d}Q' \ . \tag{3.64}$$

Due to the Dirac delta function in (3.63), the integral is only evaluated at the intersection points $\Delta Q_{1,2}$ of the parabolic potentials $V_i(Q)$ and $V_j(Q)$.

$$\Delta Q_{1,2} = \frac{c_j \Delta Q_{ij} \pm \sqrt{c_i c_j {\Delta Q_{ij}}^2 + \Delta V_{ij}(c_i - c_j)}}{c_j - c_i}. \tag{3.65}$$

Here, it has been assumed that the curvatures c_i and c_j differ. This is the case for quadratic electron-phonon coupling and two intersections of the diabatic potentials. For certain combinations of parameters (c_i, c_j, ΔV_{ij}, and ΔQ_{ij}) no intersection is obtained and the corresponding CT reaction is prohibited. Making use of the standard transformation:

$$\begin{aligned} &\int\limits_Q \delta\left(V_i(Q') - V_j(Q')\right) \mathrm{d}Q' \\ &= \int\limits_Q \sum_{l=1,2} \frac{\delta\left(Q' - Q_l)\right)}{|V_i'(Q_l) - V_j'(Q_l)|} \mathrm{d}Q' \ , \end{aligned} \tag{3.66}$$

the lineshape function can be rewritten as

$$\begin{aligned} &\xi_{ij}(c_i, c_j, \Delta Q_l, \Delta V_{ij}) \\ &= \sum_{l=1,2} \frac{1}{2}\sqrt{\frac{c_i\beta}{\pi}} \frac{\mathrm{e}^{-\beta c_i {\Delta Q_l}^2}}{|c_i \Delta Q_l - c_j(\Delta Q_l - \Delta Q_{ij})|} \ . \end{aligned} \tag{3.67}$$

The above equation is most strongly affected by the exponential factor, in which the expression $c_i {\Delta Q_l}^2$ can be identified with the energy barrier $\Delta V^\ddagger_{l,ij}$ from the minimum up to the intersection point (IP). This barrier can be expressed as

$$\begin{aligned} \Delta V^\ddagger_{l,ij}(c_i, c_j, \Delta Q_l, \Delta V_{ij}) &= \frac{c_i {\Delta Q_l}^2}{(\frac{c_i}{c_j} - 1)^2} \times \\ \times \left(1 \pm \sqrt{\frac{c_i}{c_j} + \left(\frac{c_i}{c_j} - 1\right) \frac{\Delta V_{ij}(\frac{c_i}{c_j} - 1)}{c_j {\Delta Q_l}^2}}\right)^2 \ . \end{aligned} \tag{3.68}$$

The above formula has been derived for the general case of quadratic electron-phonon coupling with differing curvatures of the parabolic potentials ($c_i \neq c_j$). By contrast, linear electron-phonon coupling is associated with the assumption of two equal curvatures ($c_i = c_j$) and yields only one intersection point at

$$\Delta Q_1 = \frac{\Delta V_{ij}/c_i + {\Delta Q_{ij}}^2}{2\Delta Q_{ij}} \ . \tag{3.69}$$

The corresponding transition barrier $\Delta V^\ddagger_{l,ij}$ is then given by

$$\Delta V^\ddagger_{l,ij}(c_i, \Delta Q_l, \Delta V_{ij}) = \left(\frac{\Delta V_{ij} + c_i {\Delta Q_l}^2}{2\sqrt{c_i}\Delta Q_l}\right)^2 \tag{3.70}$$

and the lineshape function simplifies to

$$\xi_{ij}(c_i, c_j, \Delta Q_l, \Delta V_{ij}) = \frac{1}{2}\sqrt{\frac{c_i\beta}{\pi}} \frac{\mathrm{e}^{-\beta c_i {\Delta Q_l}^2}}{|c_i \Delta Q_l|} \ . \tag{3.71}$$

In some publications [34] an alternative nomenclature has been preferred, which is based on the definitions of the Huang-Rhys factors (3.43) and (3.44) and the quantity R_i defined by

$$S_i\hbar\omega_i = R_i^2 S_j\hbar\omega_j \ . \tag{3.72}$$

In this nomenclature, the transition barrier reads

$$\Delta V^{\ddagger}_{l,ij}(S_i\hbar\omega_i, R_i, \Delta V_{ij}) = \frac{S_i\hbar\omega_i}{(R^2-1)^2}\left(1 \pm R\sqrt{\frac{S_i\hbar\omega_i + \Delta V_{ij}(R^2-1)}{S_i\hbar\omega_i}}\right)^2 \tag{3.73}$$

and the corresponding lineshape function is given by

$$\xi_{ij}(S_i\hbar\omega_i, R_i, \Delta V_{ij}) = \sqrt{\frac{\beta}{4\pi}}\sum_l \frac{R\exp(-\beta\Delta V^{\ddagger}_{l,ij})}{\sqrt{S_i\hbar\omega_i + \Delta V_{ij}(R^2-1)}} \ . \tag{3.74}$$

For linear electron-phonon coupling ($R_i = 1$), the transition barrier and the lineshape function simplify to

$$\Delta V^{\ddagger}_{l,ij}(S_i\hbar\omega_i, \Delta V_{ij}) = \frac{(\Delta V_{ij} + S_i\hbar\omega_i)^2}{4S_i\hbar\omega_i} \tag{3.75}$$

and

$$\xi_{ij}(S_i\hbar\omega_i, \Delta V_{ij}) = \sqrt{\frac{\beta}{4\pi}}\frac{\exp(-\beta\Delta V^{\ddagger}_{l,ij})}{\sqrt{S_i\hbar\omega_i}} \ , \tag{3.76}$$

respectively. At a first glance, the transition barrier (3.75) seems to depend quadratically on the driving force ΔV_{ij} for $R_i = 1$. This holds for weak electron-phonon coupling ($S_i\hbar\omega_i \ll \Delta V_{ij}$)

$$\Delta V^{\ddagger}_{l,ij}(S_i\hbar\omega_i, \Delta V_{ij}) = \frac{1}{4S_i\hbar\omega_i}\Delta V_{ij}{}^2 + \frac{1}{2}\Delta V_{ij} \tag{3.77}$$

However, this barrier dependence becomes linear for strong electron-phonon coupling ($S_i\hbar\omega_i \gg \Delta V_{ij}$), which is usually assumed in trapping models for typical conditions [34]:

$$\Delta V^{\ddagger}_{l,ij}(S_i\hbar\omega_i, \Delta V_{ij}) = \frac{1}{4}\Delta V_{ij} + \frac{1}{2}S_i\hbar\omega_i. \tag{3.78}$$

However, strong electron-phonon coupling yields a quadratic dependence for $R_i \neq 1$. This can be found by expanding equation (3.73) up to the second order:

$$\Delta V^{\ddagger}_{l,ij}(S_i\hbar\omega_i, \Delta V_{ij}) = \frac{S_i\hbar\omega_i}{(1+R_i)^2} + \frac{R}{1+R_i}\Delta V_{ij} + \frac{R_i}{4S_i\hbar\omega_i}\Delta V_{ij}. \tag{3.79}$$

III.8. Configuration Coordinate Diagram

In the previous sections, the CT reactions have been discussed on the level of diabatic potentials for a two-state molecule (defect) model. In the framework of this versatile model, the corresponding transition rates have been derived using quantities, such as ΔV^0_{ji}, V^0_i, and V^0_j, shown in configuration coordinate diagrams, such as that in Fig. 7. In the following sections, some of these quantities will be linked to defect energy levels in the band diagram, including the trap levels for the bistable defect of the four-state NMP model (see Chapter II).

We note that bistable defects have pairs of stable and metastable states, represented by a double-well in a configuration coordinate diagram. This is schematically illustrated in Fig. 8 corresponding to the four-state NMP model. Such defects show a much richer trapping dynamics compared to two-state defects. The special dynamics of the four-state NMP model originates from the fact that the charge capture or emission process proceeds through one of the metastable states $1'$ and $2'$ rather than directly between the defect states 1 and 2 or $1'$ and $2'$. The latter would result in an effective two-state model, which cannot explain the experimental TDDS data. In the following, the concept of the four-state NMP model will be discussed for an electron trap first and then extended to the case of a hole trap.

III.8.1. Electron Trap

The term "electron trap" is frequently mentioned in the context of reliability issues in electronic devices [66]. However, similar to its "hole trap" counterpart [65], this term is used ambiguously and refers to slightly distinct physical situations. For instance, in some publications, an electron trap is regarded as a defect which captures a negative charge under stress conditions and becomes neutral during recovery again ($0 \leftrightarrow -$). Most often, however, the underlying studies do not consider the alternative explanation that a positive defect can be neutralized during stress ($+ \leftrightarrow 0$). It is pointed out that such a defect would have a similar impact on the device electrostatics and is therefore nearly indistinguishable from an electron trap as originally defined. Sometimes, the electron trap is assumed to capture an electron from the substrate conduction band — even though the same behavior can be explained by hole emission into the valence band. Throughout this work, the electron trap is defined as a defect whose charge state becomes more negative. This covers all possible charge capture and emission processes, including the aforementioned variants of ($+ \leftrightarrow 0$) and ($0 \leftrightarrow -$). Furthermore, this defect can exchange charge carriers with the conduction as well as the valence band, meaning that it may undergo both electron capture or hole emission.

The dynamics of the four-state NMP model is defined by the relative energies of defect states in Fig. 8 and energy barriers connecting states 1 and $1'$ and 2 and $2'$. To provide a link between DFT calculations of defect properties at interfaces and rate calculations discussed above, we first consider a prototype Si/SiO_2 system shown in Fig. 9. It represents a periodic cell of the interface between crystalline Si and a-SiO_2 with an extra electron.

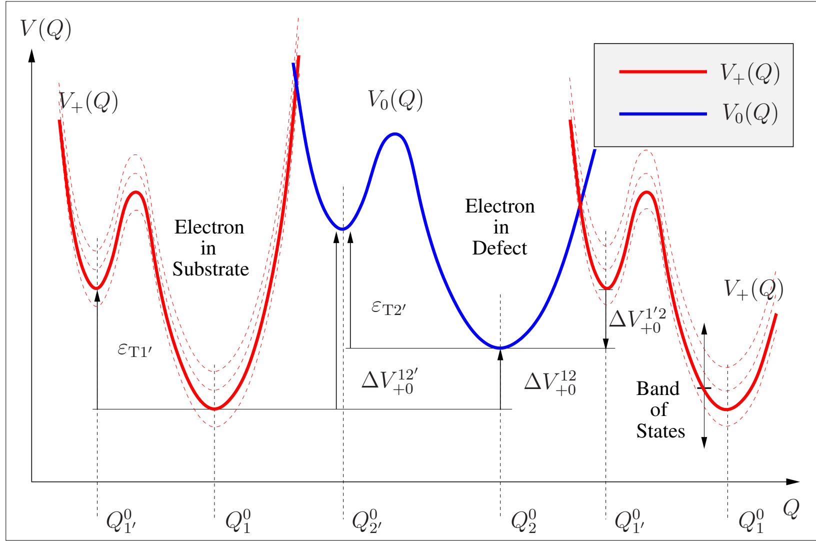


FIG. 8. Schematic configuration coordinate diagram of an electron trap which switches between bistable positive and neutral charge states. If the defect is positively charged, the exchanged electron is located in the substrate (full red curve). There, the electron can also occupy one of the energy levels E in the substrate conduction or valence band, which results in a full set of possible diabatic potentials (red dashed curves). If the defect is in its neutral charge state, the electron is captured in the trap, leaving behind a hole in the substrate. As required for the bistable defect of the four-state NMP model, the defect has one stable state (1) in the primary configuration and another (2) in the secondary configuration. As a consequence, electron capture and emission proceeds over one of the metastable states $1'$ or $2'$.

This cell is translated both along the interface and perpendicular to the interface and the total system is that of periodic interfaces. The results of such calculations obviously depend on the cell size as defects are periodically translated. In the initial state, the extra electron is delocalized in Si, as shown by the homogeneous electron density distribution in Fig. 9a. In the final state, after the electron transfer, the electron is localized on a Si dangling bond in a-SiO_2, as shown in Fig. 9b, and the electronic states in Si are perturbed by the presence of the defect. In this setup, the energy difference between the initial and final states as well as the energy barrier between the states can be calculated directly by comparing the DFT or constrained DFT total energies of the two charge configurations. Another advantage of this approach is that it directly takes account of the interaction between the defect and the interface, which results in the dependence of the tunneling rate on the relative position of the trap with respect to the interface. Such calculations, however, are very difficult, require large computer resources due to the large cell size, and therefore are very rare.

A more common approach to calculating defect energies is to consider an oxide and a substrate separately. This neglects the interaction of the defect with the interface with the substrate, but provides more flexibility for treating the defect in different charge states accurately. This is particularly important for modeling bistable defects and predicting barriers between defect configurations. The relative energies of the total system [oxide with defect + substrate] for the electron (hole) located either in the substrate or in the defect can be determined by establishing the common electronic chemical potential or Fermi level for the whole system. This is usually achieved by using the experimentally measured band offset between the substrate (e.g. Si) and the oxide (e.g. a-SiO_2). To change the charge state of the defect, an electron can be taken from or deposited to the bottom of the conduction band of Si, as schematically illustrated in Fig. 10.

To remain consistent with the definitions used for the more relevant case of a hole trap (see next section) we consider switching between positive and neutral states of a defect. The corresponding potentials for the positive ($V_+(Q)$) and neutral ($V_0(Q)$) state are depicted in the schematic configuration coordinate diagram of Fig. 8. We note that the whole system remains positively charged during the charge capture or emission event and only the location of the positive charge changes as it moves between the defect and the substrate. Furthermore, the defect features a bistability as assumed by the four-state NMP model. Therefore, the positive and the neutral charge state are represented by double-wells, having a stable (deeper) minimum in addition to a metastable (higher) energy minimum. In the positive charge state, the electron occupies a level in the band gap split from the conduction and the valence band. In order to simplify the following consideration, only a single defect state (full blue curve Fig. 8) is taken into account although the electron in the substrate can be thermally excited into different band states.

The standard way of comparing defect energies in computational materials science is to calculate the so called defect formation energies [113–115]. Most of such cal-

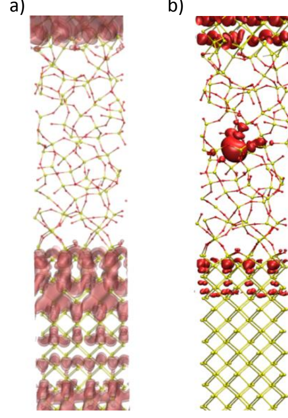


FIG. 9. Electron density distribution in a Si/SiO_2 periodic cell containing a dangling Si bond defect in the middle of the amorphous SiO_2 layer. a) The initial state of the system where an extra electron is delocalized in the Si conduction band. b) The final state of the system where the the extra electron is localized on a dangling bond.

culations are made using periodic boundary conditions and supercells capable of accommodating a defect and defect-induced distortion of the surrounding structure. The formation energy of a defect with the configuration coordinate Q in the charge state q is defined as:

$$F^q[Q] = F^q_{\mathrm{tot}}[Q] - F_{\mathrm{tot}}[bulk] - \sum_i n_i\mu_i + qE_f \ + E_{corr}, \tag{3.80}$$

where F^q_{tot} is the total energy of the system obtained from DFT calculation and $q = \ldots, -1, 0, +1, \ldots$ is the charge of the simulated system. $F_{\mathrm{tot}}[bulk]$ is the total energy for the perfect bulk crystal using an equivalent supercell. μ_i is the chemical potential of the species i with and the integer n_i is the number of host atoms or impurity atoms of type i that have been removed from ($n_i < 0$) or added to ($n_i > 0$) the supercell to form the defect. For example, formation of oxygen vacancy means removing one O atom from the supercell into the gas phase or into the Si substrate, which determines the chemical potential μ_O. E_f denotes the chemical potential of the electron reservoir (or the Fermi level position in the whole system). In our discussion E_f can be e.g. the bottom of the conduction band of the Si substrate. E_{corr} is the correction term that accounts for elastic and electrostatic interactions between defects in supercells due to periodic translation.

Defect formation energies can be used to evaluate thermodynamic transition levels (CTL) defined as the Fermi-level position for which the formation energies of charge states q_1 and q_2 are equal. This means that for the Fermi-level positions below CTL charge state q_1 is stable, while for Fermi-level positions above CTL, charge state q_2 is stable, and near CTL both charge states can coexist. We note that the Kohn-Sham eigenvalues obtained in DFT calculations are used as electronic energies in some studies. These values may give reasonable estimates in some cases but do not correctly account for the atomic relaxation during a C/E process.

The difference in defect formation energies:

$$E = F^{q+1}_{\mathrm{tot}}[Q] - F^q_{\mathrm{tot}}[Q] \tag{3.81}$$

is equivalent to

$$E = V_+(Q_1\ ) - \tilde{V}(Q_1\ ) \tag{3.82}$$

$$E = V_+(Q_{1'}) - \tilde{V}(Q_{1'})\ . \tag{3.83}$$

used in our previous discussion and in Fig. 8. The quantity $\tilde{V}(Q)$ denotes the diabatic potentials of the system, where the electron is taken out of the system consisting of the defect and the substrate and moved to Fermi level or to vacuum level.

Another important quantity with respect to CT reactions is the driving force. It has been introduced in Marcus theory and corresponds to a reaction energy in the context of adiabatic transitions. For the transition $1 \to 2'$ in the primary configuration, the driving force is given by

$$\Delta V^{12'}_{+0} = V_0(Q_{2'}) - V_+(Q_1)\ . \tag{3.84}$$

Here, the subscript '+0' of $\Delta V^{12'}_{+0}$ indicates an electron capture or hole emission process while the superscripts '$12'$' gives the states involved in the transition of the four-state NMP model. It is noted that the above expression can be traced back to typically used electronic energies

$$\Delta V^{12'}_{+0} = V_0(Q_{2'}) - V_+(Q_1) \qquad = E''_{\mathrm{t}} - E \tag{3.85}$$

using the shorthand for the thermodynamic trap level E''_{t}

$$E''_{\mathrm{t}} = V_0(Q_{2'}) - \tilde{V}(Q_1)\ . \tag{3.86}$$

The definition of the energy level E''_{t} is analog to that of the thermodynamic trap level in [56, 106, 116] and corresponds to the electronic energy that determines the

equilibrium occupancy of the defect. In an analogous manner, the driving force for the transition $1' \to 2$ can be expressed as

$$\Delta V_{+0}^{1'2} = V_0(Q_2) - V_+(Q_{1'}) = E_\mathrm{t}' - E \ , \tag{3.87}$$

where the thermodynamic trap level is defined by

$$E_\mathrm{t}' = V_0(Q_2) - \tilde{V}(Q_{1'}) \ . \tag{3.88}$$

It is emphasized here that the whole electron capture or hole emission process actually requires a transition from state 1 to 2 and is therefore determined by the energy difference

$$\Delta V_{+0}^{12} = V_0(Q_2) - V_+(Q_1) = E_\mathrm{T} - E \ . \tag{3.89}$$

The obtained quantity $E_\mathrm{T} = V_0(Q_2) - \tilde{V}(Q_1)$ is equivalent to the thermodynamic trap level for the whole process and incorporates the effect of the transition over the barrier between the states $2'$ and 2. In the following discussion we will use notation E_T for the thermodynamic trap level for the the four-state NMP model to distinguish it from E_t as a generic thermodynamic defect level.

In this context, the quantities $\varepsilon_{\mathrm{T}1'}$ and $\varepsilon_{\mathrm{T}2'}$ are worth mentioning as they relate the aforementioned trap levels E_t', E_t'', and E_T. $\varepsilon_{\mathrm{T}1'}$ corresponds to the 'relative stability' of state $1'$ with respect to a relaxation to state 1 and is given by

$$\varepsilon_{\mathrm{T}1'} = V_+(Q_{1'}) - V_+(Q_1) \ . \tag{3.90}$$

We note that this quantity must have a positive value within the four-state NMP model since the metastable state $1'$ is energetically higher than its stable counterpart 1'. The 'relative stability' of state $2'$ is defined in an analogous manner and thus reads

$$\varepsilon_{\mathrm{T}2'} = V_0(Q_{2'}) - V_0(Q_2) \ . \tag{3.91}$$

Using the above definitions (3.90) and (3.91), the trap levels E_t' and E_t'' can be expressed as

$$E_\mathrm{t}' = E_\mathrm{T} - \varepsilon_{\mathrm{T}1'} \tag{3.92}$$

$$E_\mathrm{t}'' = E_\mathrm{T} + \varepsilon_{\mathrm{T}2'} \ . \tag{3.93}$$

The above equations demonstrate that the position of E_t' and E_t'' is directly related to relative stability $\varepsilon_{\mathrm{T}1'}$ or $\varepsilon_{\mathrm{T}2'}$, respectively. For instance, E_t'' is raised by the amount of $\varepsilon_{\mathrm{T}2'}$ compared to E_T, since it requires more energy to lift the atomic system from the initial energy $V_+(Q_1)$ to the final energy $V_0(Q_{2'})$ rather than to $V_0(Q_2)$. With regard to the inverse process of electron capture or hole emission, the role of the initial and the final states are exchanged. Thus the corresponding driving forces are the inverted values of $\Delta V_{+0}^{12'}$, $\Delta V_{+0}^{1'2}$, and ΔV_{+0}^{12}.

$$\Delta V_{0+}^{2'1} = -\Delta V_{+0}^{12'} \tag{3.94}$$

$$\Delta V_{0+}^{21'} = -\Delta V_{+0}^{1'2} \tag{3.95}$$

$$\Delta V_{0+}^{21} = -\Delta V_{+0}^{12} \tag{3.96}$$

Other important quantities in the four-state NMP model are the Huang-Rhys factors and the thermal barriers. For calculating the Huang-Rhys factors, the reorganization energies $S_i\hbar\omega_i$ must be evaluated for the states $i = 1, 1', 2, 2'$ first (cf. Fig. 7).

$$S_1\hbar\omega_1 = V_+(Q_{2'}^0) - V_0(Q_{2'}^0) \tag{3.97}$$

$$S_{1'}\hbar\omega_{1'} = V_+(Q_2^0) - V_0(Q_2^0) \tag{3.98}$$

$$S_2\hbar\omega_2 = V_0(Q_{1'}^0) - V_+(Q_1^0) \tag{3.99}$$

$$S_{2'}\hbar\omega_{2'} = V_0(Q_{1'}^0) - V_+(Q_1^0) \tag{3.100}$$

The vibrational frequencies ω_i in the above equations can be calculated from the equation

$$V_i(Q_j^0) - V_j(Q_j^0) = \frac{1}{2}\omega_i^2 \Delta Q_{ij}{}^2 \ . \tag{3.101}$$

Inserting the obtained vibrational frequencies into equations (3.97)-(3.100) yields the Huang-Rhys factors.

To first order, the thermal transitions $1 \leftrightarrow 1'$ and $2 \leftrightarrow 2'$ in Fig. 8 can be described by standard transition state theory and follow an Arrhenius law

$$k_{ij} = \nu_0 \exp(-\beta\varepsilon_{ij}) \ . \tag{3.102}$$

ν_0 denotes the attempt frequency, which can be estimated by DFT calculations and is usually on the order of $10^{-13}\,\mathrm{s}$. ε_{ij} is the thermal barrier from state i to j and can be extracted from the adiabatic potentials as the energy difference from the energy minimum i up to the saddle point, which is located at $Q_{11'}^\ddagger$ or $Q_{22'}^\ddagger$ for the positive or neutral charge state, respectively. Their definitions in the four-state NMP model read:

$$\varepsilon_{11'} = V_+(Q_{11'}^\ddagger) - V_+(Q_1) \tag{3.103}$$

$$\varepsilon_{1'1} = V_+(Q_{11'}^\ddagger) - V_+(Q_{1'}) \tag{3.104}$$

$$\varepsilon_{22'} = V_0(Q_{22'}^\ddagger) - V_0(Q_2) \tag{3.105}$$

$$\varepsilon_{2'2} = V_0(Q_{22'}^\ddagger) - V_0(Q_{2'}) \tag{3.106}$$

Even though the configuration coordinate diagram provides the detailed microscopical information about the trapping process, the essence of the trapping dynamics is better represented in a band diagram. In order to translate the microscopical information to a band diagram, one has to choose one electronic reference energy first. Here, the conduction band edge is usually a good choice for an electron trap ($E = E_\mathrm{c}$). Then, the driving forces $\Delta V_{+0}^{12'}$, $\Delta V_{+0}^{1'2}$, and ΔV_{+0}^{12} can be directly plotted into the band diagram as demonstrated in Fig. 10.

III.8.2. Hole Trap

The hole trap is defined in an analogous manner to the electron trap here (see Fig. 11). Consequently, the hole trap corresponds to a defect whose charge state becomes more positive during stress conditions and returns back to its neutral charge state during recovery. Furthermore,

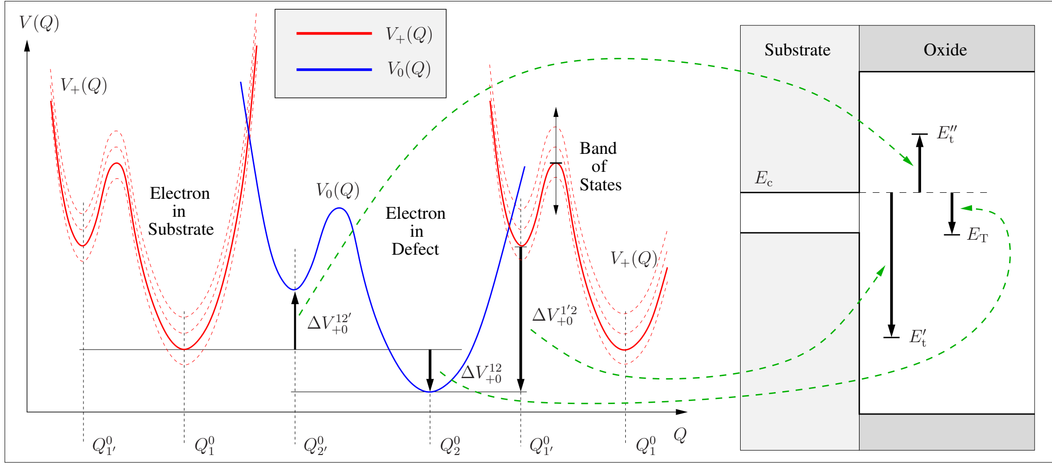


FIG. 10. Construction of the band diagram with trap level (left) from the configuration coordinate diagram (right). The conduction band edge energy is chosen as a reference level and represented by the solid red line for the diabatic potential $V_+(Q)$. According to the equations (3.85) and (3.87), the driving forces $\Delta V_{+0}^{12'}$ and ΔV_{+0}^{12} are measured from the energy $V_+(Q_1)$, respectively. For the driving force $\Delta V_{+0}^{1'2}$ (3.89), one has to keep in mind that the conduction band energy is now at $V_+(Q_{1'})$. All three energy difference can be inserted to be plotted into the the band diagram as indicated by the dashed green arrows.

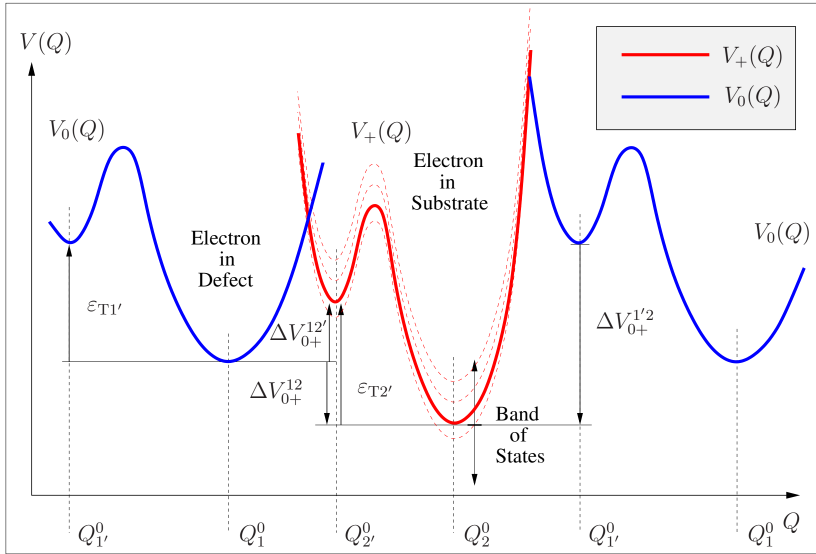


FIG. 11. The analogous defect configuration of Fig. 8 for a hole trap. In the neutral charge state, the electron is carried by the defect (blue solid line) while it is transferred to the substrate valence or conduction band in the positive charge state (red solid line). In the latter, the electron can occupy different electronic levels again.

this defect is assumed to exchange charge carriers with the conduction as well as the valence band, thereby allowing for hole capture as well as electron emission. Now the diabatic potentials $V_+(Q)$ and $V_0(Q)$ describe the atomic system where the positive charge is located in the defect or the substrate, respectively. By contrast, the potentials $\tilde{V}(Q)$ represents the reference system again, where the exchanged electron is moved to infinity. The electronic energy are then defined by the equations

$$E = V_+(Q_2) - \tilde{V}(Q_2) \tag{3.107}$$

$$E = V_+(Q_{2'}) - \tilde{V}(Q_{2'}) \ . \tag{3.108}$$

The driving forces for the transitions $(1 \to 2')$, $(1' \to 2)$, and $(1 \to 2)$ read

$$\Delta V_{0+}^{12'} = V_+(Q_{2'}) - V_0(Q_1) \tag{3.109}$$

$$\Delta V_{0+}^{1'2} = V_+(Q_2) - V_0(Q_{1'}) \tag{3.110}$$

$$\Delta V_{0+}^{12} = V_+(Q_2) - V_0(Q_1) \ . \tag{3.111}$$

Using the definitions

$$E_t'' = V_0(Q_1) - \tilde{V}(Q_{2'}) \tag{3.112}$$

$$E_t' = V_0(Q_{1'}) - \tilde{V}(Q_2) \tag{3.113}$$

$$E_T = V_0(Q_1) - \tilde{V}(Q_2) \tag{3.114}$$

they can be expressed in terms of electronic energies.

$$\Delta V_{0+}^{12'} = E - E_\mathrm{t}'' \tag{3.115}$$

$$\Delta V_{0+}^{1'2} = E - E_\mathrm{t}' \tag{3.116}$$

$$\Delta V_{0+}^{12} = E - E_\mathrm{T} \tag{3.117}$$

If one compares the driving forces $V_{+0}^{12'}$ of the electron trap and $\Delta V_{0+}^{12'}$ of the hole trap, the corresponding expressions (3.85) and (3.115) are found to have the same magnitude but differ in their signs. This is related to the fact that these transitions are actually reverse processes. Following the same arguments, one also obtains the negative driving forces for the reverse processes of (3.115) and (3.117).

$$\Delta V_{+0}^{2'1} = -\Delta V_{0+}^{12'} \tag{3.118}$$

$$\Delta V_{+0}^{21'} = -\Delta V_{0+}^{1'2} \tag{3.119}$$

$$\Delta V_{+0}^{21} = -\Delta V_{0+}^{12} \tag{3.120}$$

Using the stabilities of the states $1'$ and $2'$

$$\varepsilon_{\mathrm{T}2'} = V_+(Q_{2'}) - V_+(Q_2) \tag{3.121}$$

$$\varepsilon_{\mathrm{T}1'} = V_0(Q_{1'}) - V_0(Q_1) \ , \tag{3.122}$$

the trap levels E_t' and E_t'' can be rewritten as

$$E_\mathrm{t}' = E_\mathrm{T} + \varepsilon_{\mathrm{T}1'} \tag{3.123}$$

$$E_\mathrm{t}'' = E_\mathrm{T} - \varepsilon_{\mathrm{T}2'} \ . \tag{3.124}$$

One can recognize that the signs in the above equations differ from those in equations (3.92) and (3.93). At a first glance, this might be traced back to the different nature of the electron and the hole trap. But upon closer inspection, their configuration coordinate diagrams become similar when the roles of the positive and the neutral charge states are exchanged ($1 \leftrightarrow 2$ and $1' \leftrightarrow 2'$). For instance, the transition $1 \to 2' \to 2$ for an electron trap (see Fig. 8) corresponds to the transition $2 \to 1' \to 1$ of the hole trap (see Fig. 11). Since this correspondence is also valid for all other transitions, the electron and the hole trap can be viewed as different representations of the same defect. As such, the same defect may behave as an electron or hole trap under different operation conditions.

Hole traps are frequently observed in p-channel transistors [65] and cause charge trapping in these devices. Their trapping dynamics are dominated by hole capture and emission, meaning that the defect primarily interacts with the valence band. This case is usually discussed within the framework of the hole picture due to how intuitive it is and is outlined in the Appendix.

III.9. Interactions with the Conduction/Valence Band

In the previous section, a combination of certain points in the configuration coordinate diagram has been identified with the trap levels (E_T, E_t', E_t'') and the electronic energy E of the exchanged charge carrier in the substrate. As already pointed out before, this charge carrier may sit at a single energy level of a continuous spectrum, i.e. the conduction or the valence band. Therefore, the defect can interact with the whole conduction or valence band of the substrate and the simple formulation of the CT reactions must be extended to account for the transitions with a multitude of band states at different energies E. As these transitions occur independently of each other, they can be accounted for by summing the rate equation (3.56) over the occupied conduction or valence band states n.

$$k_{i,j} = \frac{2\pi}{\hbar}\tilde{k}^2 \sum_n f(E_n)\lambda^2(\mathbf{r}_\mathrm{d}, E_n)\,\xi_{ij} \tag{3.125}$$

The index 'n' is introduced for the carrier energy E_n in order to distinguish between the different charge carrier energies in the substrate. Furthermore, $f(E_n)$ represents the probability of finding the required charge carrier type at energy E_n. This quantity is termed energy carrier distribution function and will be denoted by f_n or $f_\mathrm{p} = 1 - f_\mathrm{n}$, depending on whether the required carrier type for the CT reaction is an electron or a hole, respectively. These charge carriers in the substrate can be assumed to be in thermal equilibrium if the source-drain bias is nearly vanishing. Then the energy-dependent occupancy $f_\mathrm{n}(E)$ of the charge carriers is given by the Fermi-Dirac distribution. If the channel of a transistor carries a significant current, the charge carriers are driven out of equilibrium and their energy profile has to be obtained by the solution of a transport formalism, such as the Boltzmann transport equation [117] or the non-equilibrium Green's function approach [118].

Since the conduction and valence band states form a continuous spectrum, the summation in (3.125) can also be transformed to an integral over a density of states, assuming the parabolic-band approximation. For the conduction band, the corresponding integral reads

$$\sum_n \to 2\Omega \int_{E_\mathrm{c}}^{\infty} \mathrm{d}E_\parallel \int_{E_\mathrm{c}}^{\infty} \mathrm{d}E_\perp D_\mathrm{2d} D_\mathrm{1d}(E_\perp) \ , \tag{3.126}$$

where $E_\perp$ and $E_\parallel$ denote the kinetic energy parallel or perpendicular to the channel interface, respectively. The factor Ω stands for the volume of the charge carriers in the substrate and the factor '2' accounts for the spin degeneracy. The one- (D_1d) and two-dimensional (D_2d) density of states [119] reads

$$D_\mathrm{1d}(E_\perp) = \frac{1}{\pi\hbar}\sqrt{\frac{m_\mathrm{eff}}{2(E_\perp - E_\mathrm{c})}} \tag{3.127}$$

$$D_\mathrm{2d} = \frac{m_\mathrm{eff}}{2\pi\hbar^2} \tag{3.128}$$

with m_eff being the effective electron mass. Using the transformation $E = E_\parallel + E_\perp$, the integral can be expressed as

$$\sum_n \to 2\Omega \int_{E_\mathrm{c}}^{\infty} \mathrm{d}E \int_{E_\mathrm{c}}^{E} \mathrm{d}E_\perp D_\mathrm{2d} D_\mathrm{1d}(E_\perp) \ . \tag{3.129}$$

With this, the corresponding electron capture rate can be written as

$$k_{+0} = \bar{k} \int_{E_c}^{\infty} \mathrm{d}E \int_{E_c}^{E} \mathrm{d}E_\perp \; D_{2\mathrm{d}} \; D_{1\mathrm{d}}(E_\perp) \; f_\mathrm{n}(E) \times \lambda^2(\mathbf{r}_\mathrm{d}, E_\perp) \; \xi_{+0}(\Delta V^0_{+0}) \quad (3.130)$$

with $\bar{k}$ being defined as

$$\bar{k} = \frac{2\pi}{\hbar} \; \tilde{k}^2 2\Omega \; . \quad (3.131)$$

The dimensionless WKB factor $\lambda(\mathbf{r}_\mathrm{d}, E_\perp)$ decreases exponentially from $E_\perp = E$ towards lower energies. As the decay of this factor dominates over the exponential dependence of the electron distribution function f_n, the integrand of the equation (3.130) has its largest contribution at $E_\perp = E$. Therefore, the dimensionless WKB factor $\lambda(\mathbf{r}_\mathrm{d}, E)$ can be evaluated at $E_\perp = E$ and equation (3.130) reduces to

$$k_{+0} = \bar{k} \int_{E_c}^{\infty} \mathrm{d}E \, D_{3\mathrm{d}}(E) \, f_\mathrm{n}(E) \times \lambda^2(\mathbf{r}_\mathrm{d}, E) \, \xi_{+0}(E, E_\mathrm{t}) \; . \quad (3.132)$$

E_t stands for E'_t or E''_t, respectively, and $D_{3\mathrm{d}}(E)$ is the three-dimensional density of states defined by

$$D_{3\mathrm{d}}(E) = \frac{m_\mathrm{eff}}{2\pi^2\hbar^3} \sqrt{2 m_\mathrm{eff}(E - E_\mathrm{c})} \; . \quad (3.133)$$

Here, the lineshape function is now expressed with respect to the carrier energy E. It is emphasized that the above equation does only apply for electron capture from the substrate conduction band but not for hole emission into the substrate valence band. Therefore, a generalized density of states is introduced, which covers the whole spectrum of states in the substrate.

$$D(E) = \frac{m_\mathrm{n}}{2\pi^2\hbar^3} \sqrt{2 m_\mathrm{n}(E - E_\mathrm{c})} + \frac{m_\mathrm{p}}{2\pi^2\hbar^3} \sqrt{2 m_\mathrm{p}(E_\mathrm{v} - E)} \; , \quad (3.134)$$

m_n and m_p are the effective charrier masses of the electrons in the conduction band and the holes in the valence band, respectively. With the generalized density of states, the electron capture rate can be rewritten as

$$k_{+0} = \bar{k} \int \mathrm{d}E \, D(E) \, f_\mathrm{n}(E) \times \lambda^2(\mathbf{r}_\mathrm{d}, E) \, \xi_{+0}(E, E_\mathrm{t}) \; . \quad (3.135)$$

It is noted that this formulation of the CT rates is also applicable to charge capture and emission with a metal gate [120]. In those materials, the density of states $D(E)$ has no bandgap and therefore can be described by one analytical expression as it has already been done in [40, 41].

The reverse processes, i.e. electron emission and hole capture, can be derived in an analogous manner to yield

$$k_{0+} = \bar{k} \int \mathrm{d}E \, D(E) \, f_\mathrm{p}(E) \times \lambda^2(\mathbf{r}_\mathrm{d}, E) \, \xi_{0+}(E, E_\mathrm{t}) \; . \quad (3.136)$$

Here, the lineshape $\xi_{+0}(E, E_\mathrm{t})$ has been replaced by $\xi_{0+}(E, E_\mathrm{t})$ and the corresponding carrier distribution function is now given by the hole occupancy function $f_\mathrm{p}(E)$.

If one considers a channel of charge carriers confined in one dimension, the summation over the energy spectrum can also be expressed as

$$\sum_n \rightarrow 2\Omega \sum_{n_\perp} \int_{E_{n_\perp}}^{\infty} \mathrm{d}E_\parallel D_{2\mathrm{d}} \; . \quad (3.137)$$

Here, $E_{n_\perp}$ corresponds to the quantized states in the direction perpendicular to the interface with the quantum number $n_\perp$ and can be obtained from the solution of a Schrödinger-Poisson solver [121]. When inserting this summation in equation (3.125), one obtains

$$k_{ij} = \bar{k} \sum_{n_\perp} \int_{E_{n_\perp}}^{\infty} \mathrm{d}E_\parallel \; D_{2\mathrm{d}} \, f_\mathrm{n}(E_{n_\perp} + E_\parallel) \times \lambda^2(\mathbf{r}_\mathrm{d}, E_{n_\perp}) \, \xi_{ij}(\Delta V^0_{ij}) \quad (3.138)$$

for electron capture ($ij \rightarrow 0+$) and electron emission ($ij \rightarrow +0$). It is noted that the prefactor $\bar{k}$ is expressed as $v_\mathrm{th}\sigma$, where v_th corresponds to the thermal velocity and σ to a cross capture section. Following the same arguments as used for the parabolic-band approximation, expressions analogous to (3.138) can be derived for hole capture and emission.

A more sophisticated approach has been presented in [122], where the energetical distribution of the charge carrier density is obtained from device simulations based on the non-equilibrium Green's functions (NEGF) method. Compared to the discrete energy levels obtained from a Schrödinger-Poisson solver, NEGF describes the energy level broadening due to diverse scattering mechanisms and provides a more accurate results, in particular when charge carriers are present. Alternatively, analytical solutions for the charge capture and emission rates have been proposed in [65, 123], where the following assumptions have been made:

- The density of states is assumed to be parabolic.
- The electron ($f_\mathrm{n}(E)$) and hole ($f_\mathrm{p}(E)$) occupancy is approximated by a Maxwell-Boltzmann distribution.
- The WKB factor is approximated by a tunneling factor of the form $\exp(-x_\mathrm{d}/x_0)$. Here, the tunneling length x_0 is defined by

$$x_0 = 2\sqrt{2 m_\mathrm{t} \Delta E}/\hbar \quad (3.139)$$

with m_t being the tunneling mass (cf. Appendix). The ΔE corresponds to the tunneling barrier, which is approximated as the conduction or the valence band offset between the substrate and the dielectric.

- In [10] the prefactor $\bar{k}$ has been interpreted as the product of the capture cross section σ and the thermal velocity v_th of the electrons or holes.
- The prefactor of the exponential terms in the lineshape function (3.71) or (3.67) is neglected.
- The CT transfer reaction must be in the strong electron-phonon coupling regime, which is explained in Fig. 12. Then the largest contribution to the lineshape function comes from those transition barriers $\Delta V_{ij}^{\ddagger}$ that are associated with charge carriers located at the band edges. The corresponding drifting forces $\Delta V_{ij}(E, E_\mathrm{t})$ in equations (3.73) or (3.75) are therefore just evaluated for electrons at the conduction band edge in the case of the electron capture/emission or holes at the valence band edge in the case of hole capture/emission.

Using the above assumptions, the derivation in [123] yields the simplified hole capture rate

$$k^{p,\mathrm{c}} = v_{\mathrm{th},p}\, \sigma_p (1+R_i)^{3/2} \exp(-x_\mathrm{d}/x_{p0})\, p \times \exp\left(-\beta \Delta V_{ij}^{\ddagger}(S_i\hbar\omega_i, R_i, \Delta V_{ij})\right) \quad (3.140)$$

where the driving force $\Delta V_{ij}(E, E_\mathrm{t})$ is evaluated for $E = E_\mathrm{v}$. Following the same derviation, the hole emission rate is obtained as

$$k^{p,\mathrm{e}} = k^{\mathrm{p,c}} \exp\left(-\beta\left(E_\mathrm{t} - E_\mathrm{f}\right)\right) \ . \quad (3.141)$$

It is emphasized that the above equation is in agreement with the well known relation

$$\frac{k_{+0}}{k_{0+}} = \frac{k_{0-}}{k_{-0}} = \exp\left(-\beta\left(E_\mathrm{t} - E_\mathrm{f}\right)\right) \ , \quad (3.142)$$

which is frequently mentioned in the context of drain current noise [10, 36, 124, 125]. Electon capture is given by

$$k^{n,\mathrm{c}} = v_{\mathrm{th},n}\, \sigma_n (1+R_i)^{3/2} \exp(-x_\mathrm{d}/x_{n0})\, n \times \exp\left(-\beta \Delta V_{ij}^{\ddagger}(S_i\hbar\omega_i, R_i, \Delta V_{ij})\right) \quad (3.143)$$

with the driving force evaluated for $E = E_\mathrm{c}$. Analogously to the relation (3.142), electron capture is linked to the reverse process of electron emission by

$$k^{n,\mathrm{e}} = k^{\mathrm{n,c}} \exp\left(-\beta\left(E_\mathrm{t} - E_\mathrm{f}\right)\right) \quad (3.144)$$

again.

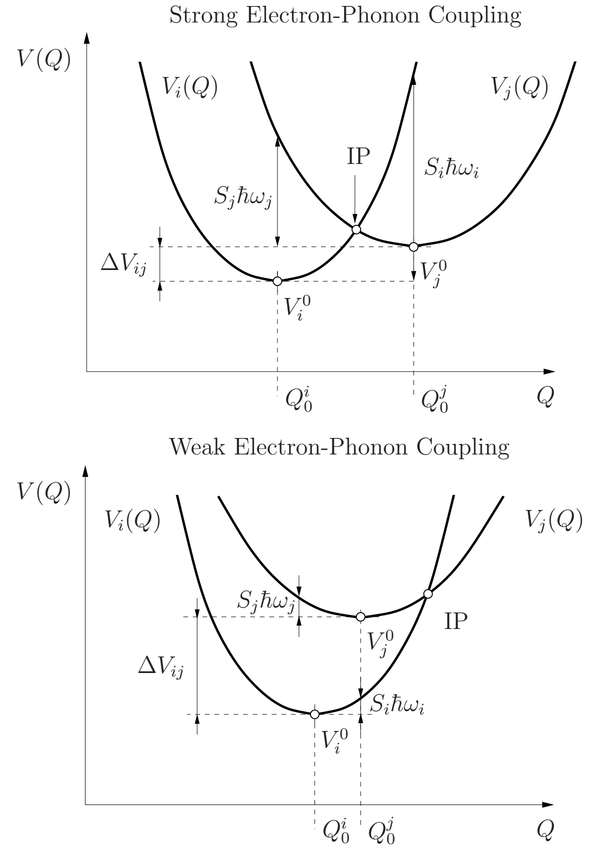


FIG. 12. Strong (top) and weak (bottom) electron-phonon coupling. In the first case the diabatic potentials are positioned such that the intersection point is situated inbetween their minima ($\Delta V_{ij} > S_i\hbar\omega_i$ and $\Delta V_{ij} > S_j\hbar\omega_j$) while in the second case one parabola lies inside the other and the intersection point is located beside the two minima ($\Delta V_{ij} < S_i\hbar\omega_i$ or $\Delta V_{ij} < S_j\hbar\omega_j$). It is noted that the term 'strong' and 'weak' electron-phonon coupling regime are also referred to as the 'normal' and the 'inverted' regime in the field of chemistry. In addition, by changing the gate bias a defect may go from the weak to the strong coupling regime or vice versa.

IV. STOCHASTIC DYNAMICS OF ELECTRON TRANSFER

In the previous sections, we have discussed CT reactions and pure thermal transitions along with their underlying theory. However, the complex trapping dynamics seen experimentally are actually caused by a combination of those processes and involve several meta-/stable states as pointed out in Sec. II. These dynamics is stochastic and can be correctly treated using homogeneous continuous-time Markov chain theory [126]. This theory rests upon the assumption that the future transitions between states do not depend on the past of the investigated system. This assumption is justified as long as the defect undergoes structural relaxation after each transition. Thereby, the defect interacts with its environment and loses the memory of its past. In fact, this is assumed to be the case for both pure thermal and NMP transitions except for more specialized theories,

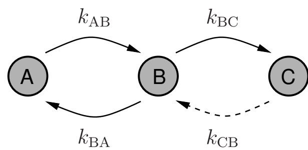


FIG. 13. The state diagram for a two-step process from the state A to C. The first passage time of such a process is calculated by equation (4.2). Note that the transition rate k_{CB}, indicated by the dashed arrow, does not enter this equation.

such as recombination-enhanced defect reaction [127]. The time evolution of such a defect system is described by a first-order differential equation system, known as Master equation.

$$\partial_t \pi_i(t) = \sum_{j\neq i} \pi_j(t) k_{ji} - \sum_{i\neq j} \pi_i(t) k_{ij} \tag{4.1}$$

Here, $\pi_i(t)$ is the time-dependent probability that the defect is in state i and k_{ij} denotes the transition rate from state i to state j. The above stochastic differential equation can be treated by the Gillespie algorithm [128], which allows one to generate numerous statistically correct solutions at computationally feasible costs. Each of these solutions corresponds to one possible realization where many of them must be computed to gather statistically meaningful information. If a sufficiently large number of such realizations has been computed, one can analyze the trapping behavior of single defects. For instance, one can determine the expectation value for the actual transition times between each combination of states — and therefore also the capture and emission times between the states 1 and 2 in the four-state NMP model.

As the above Master equation describes a first-order process, the occupation probabilities can be averaged, thereby becoming occupancies. The resulting rate equations are of the same form as the above Master equation and can be solved numerically as a partial differential equation. In this way, the computation of a vast number of realizations can be avoided and the rate equation for one defect can be solved efficiently. Thereby, the calculation of a large number of defects becomes computationally feasible. Since these defects can have different properties, this method also accounts for variations in the defect properties and thus allows to compute their distributions. These variations may arise from the defect properties or the amorphous defect environments but can also be attributed to random dopant fluctuations, which have increasingly attracted scientific interest during the last several years [10, 129].

For a comparison to the single-defect data of TDDS measurements, one is primarily interested in the transition times between stable states. The metastable states are only occupied temporarily and therefore barely observed in experiments. However, they become relevant for the overall gate bias and temperature dependence of two-step processes. The transition rates between stable states can be obtained from first-passage times. This quantity corresponds to the mean time it takes the considered system to first arrive at state C, provided that it was in state A but not in state B at the beginning (cf. Fig. 13). The corresponding transition time reads [34]

$$\begin{aligned}\tau_{AC} &= \frac{k_{\mathrm{AB}} + k_{\mathrm{BC}} + k_{\mathrm{BA}}}{k_{\mathrm{AB}} k_{\mathrm{BC}}} \\ &= \frac{1}{k_{\mathrm{AB}}} + \frac{1}{k_{\mathrm{BC}}} + \frac{1}{k_{\mathrm{BC}}}\frac{k_{\mathrm{BA}}}{k_{\mathrm{AB}}} .\end{aligned} \tag{4.2}$$

In the multi-state model, the set of four states allows for different transition pathways (see Fig. 3). The corresponding first-passage times are listed below

$$\tau_{\mathrm{c}}^{2'} = \frac{1}{k_{12'}} + \frac{1}{k_{2'2}} + \frac{1}{k_{2'2}}\frac{k_{2'1}}{k_{12'}} \tag{4.3}$$

$$\tau_{\mathrm{c}}^{1'} = \frac{1}{k_{11'}} + \frac{1}{k_{1'2}} + \frac{1}{k_{1'2}}\frac{k_{1'1}}{k_{11'}} \tag{4.4}$$

$$\tau_{\mathrm{e}}^{2'} = \frac{1}{k_{22'}} + \frac{1}{k_{2'1}} + \frac{1}{k_{2'1}}\frac{k_{2'2}}{k_{22'}} \tag{4.5}$$

$$\tau_{\mathrm{c}}^{1'} = \frac{1}{k_{21'}} + \frac{1}{k_{1'1}} + \frac{1}{k_{1'1}}\frac{k_{1'2}}{k_{21'}} . \tag{4.6}$$

The rates $k_{11'}$, $k_{1'1}$, $k_{22'}$, and $k_{2'2}$ corresponds to the adiabatic transitions following equation (3.9). Here, these rates are expressed as

$$k_{ij} = \nu_0 \exp(-\beta \varepsilon_{ij}) \ , \tag{4.7}$$

where ε_{ij} is the thermal barrier between from state i to state j and the attempt frequency ν_0 is the order of $10^{13}\,\mathrm{s}^{-1}$.

As pointed out in Sec. II, these pathways have been identified as those transitions visible in stimulated charge trapping and illustrated in the state diagram of Fig. 3. Here, the capture and emission times are the response to the alternate application of a high- and a low-level gate bias as in stimulated charge trapping. However, the same transitions are also encountered for most of the investigated defects at a constant gate bias level and therefore observed as drain current RTN [129]. As such, the general description based on a four-state trap already covers the most important charge trapping phenomena for the relevant use-conditions of microelectronic transistors. Besides, the four-state NMP model allows for many other transitions which might become relevant for charge trapping phenomena. Some of them have not been investigated experimentally in full detail while others may not be visible due to a too low resolution of the measurement equipment. Nevertheless, there are some further noise phenomena which could be explained within the framework of the four-state NMP model and will be discussed in the following.

One of them is anomalous RTN, which was discovered in the early studies of Uren [10]. They observed electron traps, which repeatedly produced noise for random time intervals. During the interruptions of the noise signal, the defects were found to remain negatively charged and did not generate RTN noise. Within the four-state NMP model, the noise signal is generated by NMP transitions between the states of the secondary configuration

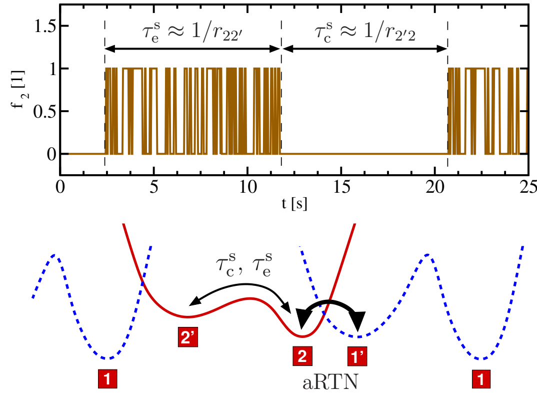


FIG. 14. **Top**: The hole occupancy f_2 during aRTN with the gate bias held at a constant level. **Bottom**: Configuration coordinate diagram for an aRTN defect. Since this defect is a hole trap, the red solid and the blue dashed line correspond to the positive and neutral charge state, respectively. The double-sided thick arrow is associated with aRTN while the thin one represents the transitions into and out of the metastable state $2'$.

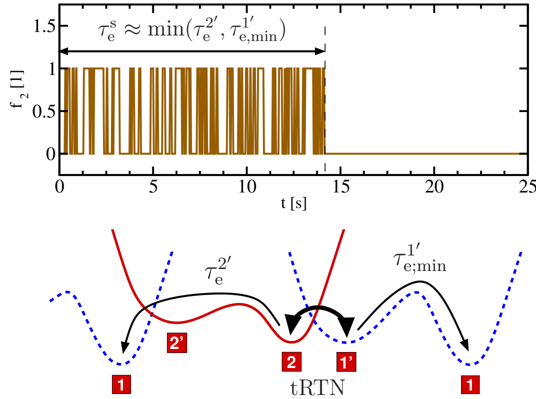


FIG. 15. **Top**: The same as in Fig. 14 but for tRTN. At $t = 0$ the stress voltage has been removed and the defect is in its positive state 2. After a time τ_e^s the defect ceases to produce noise. **Bottom**: Configuration coordinate diagram for a tRTN defect. The double-sided thick arrow indicates the fast NMP transitions between the states 2 and $1'$ related to the occurrence of noise. The possibilities to escape from these states are shown by the thin arrows.

(see Fig. 14). Therefore, the corresponding capture and emission time constants are given by the direct NMP transition rates $k_{1'2}$ and $k_{21'}$, respectively. The recurrent pauses of the noise signal are initiated by transitions from the stable state 2 to the metastable state $2'$. During the time of the interruption, the defect dwells in this state and no NMP transition can take place. Here it has been implicitly assumed that the NMP transition $2' \to 1$ occurs on larger time scales than the return to the state 2 via the transition $2' \to 2$. The emission time constant τ_e^s in Fig. 14 defines the mean time interval during which noise can be observed. Its value is given by the inverse of the transition rate $1/r_{22'}$. The capture time constant $\tau_c^s = 1/r_{2'2}$ corresponds to the mean time interval until the next noise period starts.

Another noise phenomenon [65] has been observed in TDDS measurements, where the drain current displayed RTN only after a high-level gate bias had been previously applied. Microscopically, the noise at low-level gate bias can be linked to defects which continuously capture and emit charge carriers. In the measurements, the RTN signal was observed to disappear after a random amount of time. However, it did not reoccur during the remaining measurement time in contrast to anomalous RTN. The disappearing noise signal was ascribed to a transition into a neutral charge state, in which the defects remain. As this RTN behavior ends after a random amount of time, this phenomenon is referred to as temporary RTN in TDDS experiments. Within the four-state NMP model, the generated noise originates from defects switching back and forth between states 2 and $1'$. Since the associated NMP transitions $2 \leftrightarrow 1'$ do not involve any intermediate states, the capture and emission times are given by the inverse of the transition rates $k_{1'2}$ and $k_{21'}$, respectively. As illustrated in Fig. 15, the noise states $1'$ or 2 can be reached by the transition $1 \leftrightarrow 2' \to 2$ at the high gate bias level. The other direct pathway $1 \to 1'$ into one of the states $1'$ or 2 is assumed to be suppressed by a large thermal barrier. At the low gate bias level, the state 1 is thermodynamically favored due to its energetically lower position compared to the states 2 and $1'$. When the defect returns to its initial state 1, the RTN signal disappears with a time constant of τ_e^s. The corresponding transition could be either $2 \to 2' \to 1$ or $1' \to 1$ with a time constant of $\tau_e^{2'}$ or $\tau_{e;min}^{1'}$, respectively (cf. Fig. 15). The actual termination of the noise signal after a time period of τ_e^s is determined by the minimum of these time constants. It is noted that the NMP barriers $2 \to 1'$ and $1' \to 2$ must not be too large since otherwise trapping events will not be observed within a practically relevant time period.

One should keep in mind that defects showing an aRTN behavior can also be responsible for tRTN. When a high-level gate bias is applied, the defects are forced into one of the states 2 and $1'$ where they produce an RTN signal. As in aRTN, they undergo a transition to the metastable state $2'$, accompanied by the end of the noise signal. However, this special sort of defects is characterized by a large capture time constant $\tau_e^s = 1/r_{2'2}$, which is larger than the typical measurement window of one TDDS cycle. As a consequence, the next transition back to the state 2 and the subsequent noise period are shifted out of the experimental time window of TDDS and will not be recorded during the measurement run. According to this explanation, tRTN can also be explained as a stimulated variant of aRTN.

V. DEFECT CANDIDATES

The bistability in both charge states is the key feature of the four-state NMP model and is thus used as the main criterion in the search for possible defect candidates. Together with a thermodynamic level somewhere

in the middle of the bandgap of SiO_2, this criterion turns out to be quite restrictive, making a large number of defects known in the literature unsuitable. The first candidate suggested for the four-state NMP model was the oxygen vacancy, which is the most prominent defect known already from the HDL model. Its bistability was predicted by a large number of independent theoretical investigations [49, 51–53, 130] and motivated for the HDL model [69, 131] to explain the switching characteristics observed after irradiation. However, DFT simulations [74, 132] have revealed that this defect does not show the trapping behavior which has been found for the defects investigated by TDDS because its trap level lies too deep in the SiO_2 bandgap.

Other defect candidates are hydrogen-complexed defects, as evidenced by ESR measurements [70, 133]. One of them is the hydrogen bridge, which was studied in the context of gate leakage currents in transistors by Bloechl *et al.* [7] but has not been related to charge trapping phenomena, such as drain current RTN and BTI, so far. In addition to the hydrogen bridge, a newly discovered/novel defect termed the hydroxyl E' center was proposed fifteen years ago by Balk [134] and recently confirmed as a promising candidate by DFT investigations [135, 136]. This defect has rarely been studied theoretically since it is not stable in crystalline SiO_2, which is often used in DFT studies due to its well-known structure and reduced computational costs [137, 138]. This underlines the necessity to perform the atomistic DFT simulations on realistic amorphous host structures, where variations in the bond length and angles allow for new defect structures. In the following, the oxygen vacancy (OV), the hydrogen bridge (HB) as well as the hydroxyl E' (HE) center were chosen for our DFT investigation as they could be consistent with the trapping behavior seen in TDDS.

V.1. Details of the DFT Calculations

For efficiency reasons, a combination of classical force-field molecular dynamics (MD) and a subsequent DFT optimization was employed to generate amorphous SiO_2 host structures. The MD simulations were based on the ReaxFF force-fields, parametrized to reproduce the properties of various silicon and silica polymorphs [139, 140]. The initial SiO_2 structures contained 216 atoms and underwent a melt and quench procedure [141] in order to produce amorphous SiO_2. In the subsequent step, DFT simulations were employed to geometrically optimize these structures. These simulations were carried out using the CP2K code [142] with the non-local functional PBE0 TC LRC [143], which predicts accurate bandgaps and localized states therein. The geometrical optimization used the Broyden-Fletcher-Goldfarb-Shannon (BFGS) method to minimize the forces on the atoms below ($2.3 \times 10^{-2}\,\mathrm{eV\,Å^{-1}}$) as well as the total energies of the defect structures. The above procedure yielded amorphous structures, which consisted of a defect-free continuum random network of SiO_4 tetrahedra and had an averaged density of $2.16\,\mathrm{g\,cm^{-3}}$. Geometrical analysis of the obtained structures agreed well with other previous calculations [144] and experimental neutron-diffraction data [145]. Due to the applied periodic boundary conditions, charged supercells were calculated by introducing a homogeneous compensating background charge. The barrier height between two different configurations was determined using the climbing-image nudged-elastic-band method (CI-NEB) [146] with a spring constant of $2\,\mathrm{eV\,Å^{-2}}$.

Dataset	2 states	3 states	4 states
HB	61	9	79
OV	63	53	42
HE	51	—	49
HEbreak	159	—	—
HEstick	67	—	—

TABLE I. Number of investigated defect structures for the hydrogen bridge (HB), the oxygen vacancy (OV), and hydroxyl E' centers (HE). The generated defects remained stable in two, three, or four configurations, which may be identified with the states of four-state NMP model.

In order to gather sufficiently accurate statistics, a large number of defects have been investigated (see Tab. I). This array included defect structures with 2, 3, or 4 stable configurations, which correspond to the states in the four-state NMP model. The relative numbers of those defects are not related to the corresponding defect concentrations in real oxide structures. Rather, these numbers were chosen so that a statistical interpretation of the properties for each investigated defect type is justified (except for the 3-state HE). For this purpose, the more complex defect structures with 3 or 4 stable configurations were created at defect sites where the defects are likely to have bistable defect configurations. The four obtained defect structures were grouped into configurations A and B, where one of them corresponds to the primary configuration and the other to the secondary configuration of the four-state NMP model. The defect configurations were also distinguished by their charge states and therefore denoted as A^0, A^+, B^0, or B^+. However, these configurations do not directly correspond to the states of the four-state NMP model since the latter are differentiated according to their relative energies.

V.2. Oxygen Vacancy

In stoichiometric SiO_2, two silicon atoms are always connected by one bridging oxygen atom. If the latter is removed, the two neighboring silicon atoms establish a dimer bond, illustrated as configuration A^0 in Fig. 16. DFT simulations predict a bond length of approximately 2.5 Å in crystalline SiO_2 [52, 116] while the bond varies in the range of $2.3 - 2.7$ Å [49, 147] in amorphous SiO_2 host

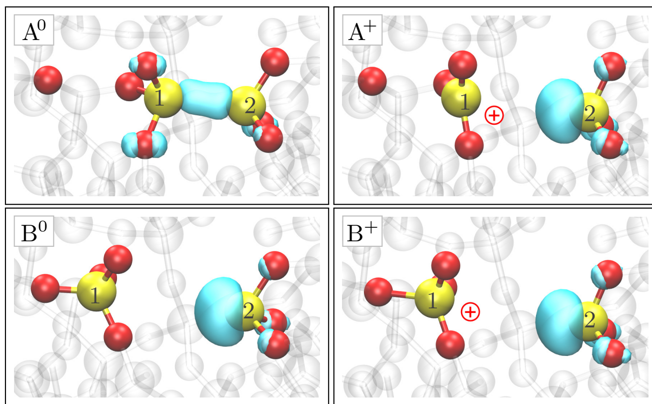


FIG. 16. Oxygen vacancy in the configurations A^0 (top left), A^+ (top right), B^0 (bottom left), and B^+ (bottom right). In the neutral charge state, the oxygen vacancy is most stable in the configuration A^0, which is characterized by a dimer bond. When this defect is positively charged (A^+), this dimer bond is either broken or stretched due to the positive charge (not shown here). The oxygen vacancy can also undergo a defect transformation called puckering, where one Si atom carries a dangling bond while the other Si atom bonds to a network O atom in the back. The corresponding configurations are denoted as B^0 and B^+ for the neutral and the positive charge state, respectively.

structures. Nicklaw [130] extended his defect calculations to highly strained oxygen vacancies with bond lengths up to 3.2 Å.

The positively charged counterpart of the oxygen vacancy is labeled configuration A^+, which is identified with the E'_δ center in the EPR measurements. The missing negative charge within its bond results in a repulsive force between the two positively charged silicon atoms and in a stretching or a breakage of the Si-Si bond. In the crystalline SiO_2 reference, the Si-Si bond of the E'_δ center extends from 2.5 Å to 3.0 Å [52, 116] upon hole capture. In $a-SiO_2$, this kind of weak bond experiences large tensile and compressive forces due to the amorphous nature of the host material. Therefore, the corresponding bond lengths were found to be widely distributed within a range of 3.0 − 4.4 Å [130], consistent with the values given in [52] and [51].

The E'_δ center can also transform to an E'_γ center, which corresponds to the bistable partner B^+ of the oxygen vacancy [51, 52, 69, 131]. Starting from the configuration of the E'_δ center, one side of the defect undergoes a transformation called ‘back-projection’ or ‘puckering’. During this process, the dimer bond is broken and one of the silicon atoms moves through the plane defined by its three oxygen neighbors. This defect structure is referred to as the back-projected configuration [49]. If, however, the Si atom is stabilized in this position via formation of a back bond to a nearby oxygen atom, the resulting configuration is termed puckered [51]. On the other side, the defect transformation leaves behind a dangling Si bond, which can carry up to two electrons and produces an EPR signal depending on its occupation [148]. In crystalline SiO_2, the puckering transformation proceeds over a small thermal barrier of 0.4 eV, where the final ‘puckered’ configuration B^+ are less stable by 0.3 eV [52]. The neutral charge state of the E'_γ center is susceptible to return back to its oxygen vacancy configuration [51]. In crystalline SiO_2, however, the DFT calculations of Mysovsky *et al.* [53] predicts that the E'_γ center is more stable via a puckering transformation over a small thermal barrier of 0.2 eV. The corresponding configuration B^0 of the E'_γ center is shown in Fig. 16.

V.3. Hydrogen Bridge

The hydrogen bridge [56, 116, 149–152] can be thought of as a $Si-Si$ dimer bond decorated by a hydrogen atom. For instance, this defect may be formed by the exothermic reaction of atomic hydrogen and an oxygen vacancy [135]. Both constituents of the hydrogen bridge, namely the oxygen vacancy as well as the hydrogen atom, were experimentally found in abundance in amorphous SiO_2: The former was experimentally confirmed by ESR studies [1, 2] but also theoretically predicted in larger amounts close to the Si-SiO_2 interface as suboxides [153]. The second constituent (hydrogen in its atomic, molecular, or a bound form) is known to exist in large background concentrations of up to $10^{19}\,cm^{-3}$, even within dry oxides of MOS transistors [70]. For instance, it was observed using nuclear reaction analysis [154, 155]. Due to the large availability of both constituents, the hydrogen bridge became one of the prime suspects in our defect search.

In order to explore the above hydrogen reaction in greater detail [135], an oxygen vacancy was created as described Chapter V.2 and then a hydrogen atom was placed close to several oxygen vacancy sites. This was done for 144 sites within our amorphous SiO_2 structures in order to collect sufficiently large statistics as required for amorphous materials. Geometrical optimization of these structures always resulted in hydrogen bridges, which are more stable by 2.76 eV on average than their separated constituents.

In the configuration A^0 (see Fig. 17), the hydrogen atoms establish a strong bond to one Si atom with a bonding distance of 1.47 Å on average. By contrast, the interactions with the other Si atoms are of a non-bonding character. These interactions are therefore strongly influenced by the amorphous environment, resulting in a wide distribution of the corresponding $Si-H$ bond lengths with a range between 1.74 Å and 3.13 Å. In the positive charge state (configuration A^+), however, the defect sometimes forms a common $Si-H-Si$ bond, with the two bond electrons shared between the three defect atoms of this chain. Since the two Si atoms of this defect point towards each other, the configurations A^0 and A^+ are often referred to as closed hydrogen bridges. In addition, there is also a broken hydrogen bridge, which is formed when the Si atom with the dangling bond moves through the plane of its three oxygen neighbors. Just like oxygen vacancy, this configuration may also be stabilized by a weak

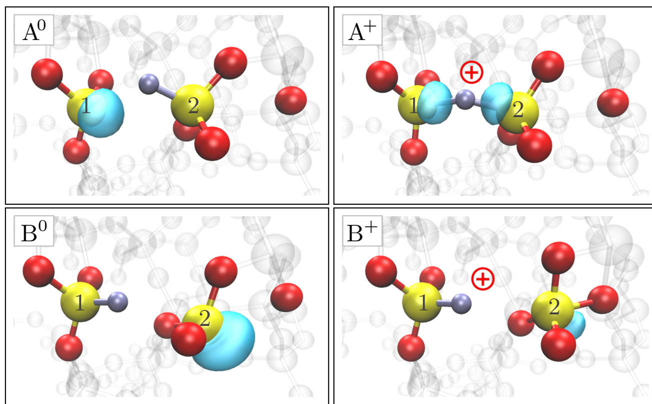


FIG. 17. The same as Fig. 16 but for the configurations of the hydrogen bridge. In the initial neutral configuration A^0, the hydrogen atom bonds to the Si atom on the right-hand side but only weakly interacts with the left Si atom. Upon hole capture, the defect adopts the configuration A^+ where the central hydrogen atom establishes a bond to both Si atoms. A possible secondary configuration is associated with the puckering of one Si atom, which points towards (configuration B^+) or even bonds to a back-oxygen (configuration B^0).

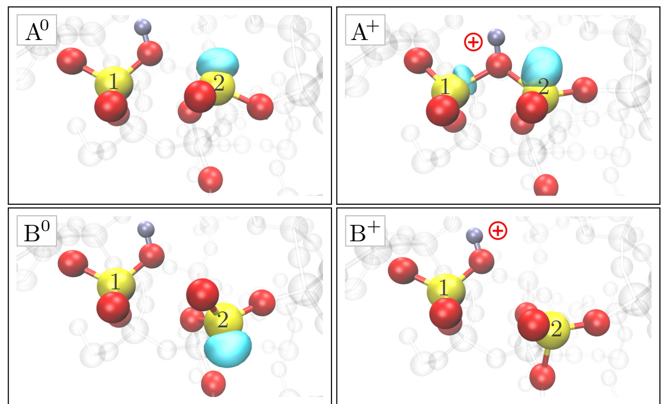


FIG. 18. The same as Fig. 16 but for the configurations of the HE center. The initial neutral configuration A^0 consists of a hydroxyl group and a Si dangling bond. In the positive charge state, they recombine to a bridging oxygen atom with an attached proton (configuration A^+). The secondary configurations involve the puckering of the Si atom again. Just as in the case of the hydrogen bridge, the puckered Si atom is directed towards a bond to the back-oxygen (configuration B^0) in the neutral charge state while it forms a bond in the positive charge state (configuration B^+).

bond to another nearby oxygen atom. This configuration remains stable for the neutral as well as the positive charge states (configurations B^0 and B^+, respectively). As a result, the hydrogen bridge exhibits the required bistability and is thus considered a potential hole trap. It is also noted here that the hydrogen bridge also has a stable configuration in its negative charge state, making it predominantly an amphoteric trap. Therefore, the hydrogen bridge can also capture and emit negative charges and may act as an electron trap, as observed in TDDS measurements on nMOS transistors [66].

V.4. Hydroxyl E' Center

The presence of HE centers in MOS transistors is a direct result of the amorphous nature of SiO_2, which exhibits a wide distribution of bond lengths and angles. Especially in regions of tensile strain in the $a{-}SiO_2$, the $Si-O$ bond lengths can sometimes strongly deviate from the crystalline equilibrium values of 1.61 Å, making their corresponding bonds chemically reactive [135]. These strained bonds were the subject of several experimental studies but have also recently been investigated with regard to possible reactions with hydrogen. It has been discovered that neutral hydrogen atoms preferentially attacks strained $Si - O$ bonds, forming a defect referred to as HE center. This defect consists of two moieties: a three-fold coordinated Si atom, which carries a dangling bond, and a hydroxyl group, which is attached/bound to the silica network (see Fig. 18). Interestingly, the HE center is energetically more stable than the hydrogen atom in its interstitial position by 0.8 eV on average. Furthermore, this defect only occurs at strained Si-O bond sites with a bond length larger than 1.65 Å. This condition is met by approximately 2% of all $Si - O$ bonds in amorphous SiO_2 so that amorphous SiO_2 has a concentration of $1.7 \times 10^{21}\,\mathrm{cm}^{-3}$ possible defect sites. For these reasons, the HE center is assumed to exist as a stable defect present in significant concentrations and must consequently be considered as an intrinsic defect of amorphous SiO_2.

In addition to the neutral charge state (configuration A^0), this defect is also stable as a positive defect, whose configuration (A^+) is strongly distorted as shown in Fig. 18. There, the oxygen atom from the hydroxyl group rebonds to the Si atom with the dangling bond so that the resulting familiar structure is that of a proton bonded to a bridging oxygen atom. The HE center also stays stable in its negative charge state according to DFT calculations in [135]. It can, therefore, be classified as an amphoteric trap or a negative-U center at 30 or 70 percent of all possible sites in $a{-}SiO_2$ [135]. The former can be present at least in a negative, the neutral, and a positive charge state depending on the position of the Fermi level in the semiconductor, while the latter directly switches between a positive and a negative charge state. However, these categories refer to the equilibrium rather than the dynamic properties of those defects. For instance, the negative charge states can be inaccessible during the experimentally relevant timescales due to excessively large NMP transition barriers and the trapping dynamics would then be dominated by NMP transitions between the positive and the neutral charge state.

In addition to the above configurations, this defect also features a back-projected configuration, in which the Si dangling bond is displaced through the plane of its three oxygen neighbors and points away from the hydroxyl group. This configuration is stable in the positive as

well as in the neutral charge state (configuration B^+ and B^0, respectively) for the vast majority of defects if the back-projected configuration is stabilized by a bond of the silicon atom with a nearby network oxygen atom. As such, the HE center shows a bistability in two charge states.

The defect structures of the hydrogen bridges were created by selecting different sites within $a{-}SiO_2$, where the hydrogen atom was placed close to an oxygen atom with an extremely long Si-O bond [156]. This selection criterion resulted in HE centers (0.14%) which also showed a back-projected configuration. Our comprehensive study on HE centers also included datasets in which the defects were created without the aforementioned selection criterion. There, the hydrogen atom was found to either stick to the closest network oxygen atom. Analogously to [135], the resulting defect is denoted as $\{SiO_4/H\}$, where the curly brackets indicate the omission of the selection criterion. Alternatively, the hydrogen atom can also break one Si-O bond and form a hydroxyl group, which is referred to as HE. It is noted that the datasets $\{SiO_4/H\}$ and $\{HE\}$ show differently distributed properties since the HE centers with at least four states represent a small subgroup due to the selection criterion.

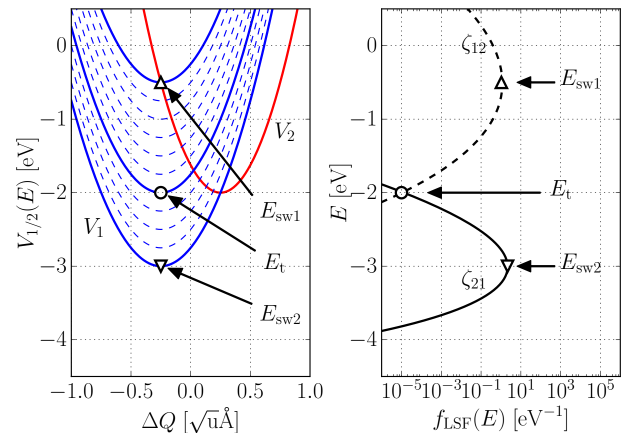


FIG. 19. Schematic configuration coordinate diagram (left) for electron capture. During the CT reaction, the electron with the energy E is in state 1 and undergoes a transition to state 2. The involved states are represented by the initial potentials $V_1(E)$ and the final potential V_2. The thermodynamic trap level E_t corresponds to the energy E at which the energy minima of the initial and the final potential are equal. In the plots of the lineshape functions (right), this coincides with the point where $\zeta_{12}(E)$ and $\zeta_{21}(E)$ intersect. The switching trap level E_{sw1} is given by the energy E whose corresponding potential $V_1(E)$ intersects with the potential V_2 in its minimum. At this energy level, the vibrational transition proceeds at the fastest pace since no thermal excitation is required. Similarly, the lineshape function ζ_{21} peaks at the switching trap level E_{sw2} where the potential $V_1(E)$ cuts the potential V_2 in its minimum.

VI. PARAMETER EXTRACTION FROM DFT

In the previous section, we have presented several promising defect candidates which feature a bistability as required by the four-state NMP model. In this section, they will be evaluated primarily based on their trap levels calculated by DFT. Furthermore, other important model parameters, such as the relaxation energies ($S_i\hbar\omega_i$, $S_j\hbar\omega_j$) and the displacement (ΔQ_{ij}), will be studied as they have a strong impact on the NMP transition rates.

For this purpose, the distributions of these quantities will be extracted from our DFT simulations for three defect candidates. Based on these data, realistic ranges will be determined for the aforementioned model parameters. The parameter ranges are regarded as benchmarks and will be used to evaluate fits of the model to experimental TDDS data. Furthermore, it will be demonstrated that nuclear tunneling can strongly affect the NMP transition rates under certain conditions (low temperatures, specific values of the model parameters $S_i\hbar\omega_i$, $S_j\hbar\omega_j$ and ΔQ_{ij}). As such, nuclear tunneling will be investigated with respect to its impact on the magnitude as well as the gate bias and temperature dependence of the NMP transition rates. In particular, we will give estimates for the error made by assuming the classical approximation for 'typical' ($T \in [300\,\mathrm{K}, 500\,\mathrm{K}]$ and $|E_{ox}| < 10\,\mathrm{MV/cm}$) and 'extreme' ($T \in [300\,\mathrm{K}, 650\,\mathrm{K}]$ and $|E_{ox}| \lesssim 10\,\mathrm{MV/cm}$) operation conditions of microelectronic transistors.

VI.1. Trap levels

A crucial parameter of any defect model is the thermodynamic trap level since it determines whether a defect can be charged and discharged under certain bias conditions. This is explained for the electron capture in the configuration coordinate diagram of Fig. 19. Here, the state 1 corresponds to the case where the electron sits in one of the band states, represented by the set of parabolas $V_1(E)$. In the state 2, however, the electron is located in the defect, represented by the parabolic potential V_2. If the potential $V_1(E)$ is located above V_2, the electron is preferably in the state 2 and thus in the defect. Otherwise, the electron is in state 1 and occupies one of the band states. As such, the location of the electron is related to the driving force

$$\Delta V^{12}_{+0} = E - E_t \tag{6.1}$$

(cf. equations (3.85) and (3.87)) and depends on the energy difference $E - E_t$. Recall that the trapping dynamics involve not just one but a multitude of band states E. As mentioned in Sec. III, these states act as an electron reservoir in CT reactions and their occupancy is related to the substrate Fermi level E_f. As a consequence, the occupation of the defect is ultimately given by the Fermi level. This level can be associated with

the parabola $V_1(E_f)$, which corresponds to the highest of the occupied band states in the band energy diagram. If this parabola $V_1(E_f)$ is located above V_2, the defect is occupied, and otherwise empty. At the transition between these cases, the parabolas $V_1(E_f)$ and V_2 are at the same height ($E'_t \approx E_f$) and the defect and the band states E at the Fermi level are equally occupied. For example, this is the case for drain current noise where the electron capture and emission balances each other and the corresponding capture and emission times are of the same order of magnitude. However, the noise signal often remains unnoticed as it can only be resolved experimentally if the time constants fall within the measurement window.

In addition, the thermodynamic trap level also determines whether stimulated charge trapping can occur for the gate bias repeatedly switching between two voltage levels. If the thermodynamic trap level is situated below the Fermi level, electron capture (hole emission) dominates over electron emission (hole capture). Consequently, the defect adopts its "more negative" charge state in thermal equilibrium. By contrast, if the thermodynamic trap level lies above the Fermi level, the roles of capture and emission are reversed and the defect changes to its "more positive" charge state. Therefore, stimulated charge trapping requires the thermodynamic trap level to pass the Fermi level during a bias sweep. As a consequence, only those defects which have their thermodynamic trap levels shifted above and below the Fermi level during a whole period, are visible in TDDS experiments. This condition is met for those defects located in the active region, shown in Fig. 20 for a pMOSFET [157].

It is emphasized here that the thermodynamic trap levels should not be confused with switching trap levels [7]. The latter are associated with zero-barrier transitions and are related to the maxima of the lineshape function for hole capture and emission (see Fig. 19). However, the switching trap levels have usually no particular relevance at normal bias conditions during charge trapping and will therefore not be addressed here.

In the four-state NMP model, the primary and secondary configuration have their own thermodynamic trap levels, E''_t and E'_t, respectively which can be extracted from the DFT simulations using equations (3.112) and (3.113). These trap levels are based on the classification of stable (1, 2) and metastable (1′, 2′) states. However, this definition does not necessarily agree with the 'configurations' (A^0, B^0, A^+, B^+), which are related to the atomic structure of the defects. For this reason, the configurations (A,B) of the defect candidates must be identified with the states of the four-state NMP model. Since the shape of the neutral and the positive potentials vary from defect to defect, a certain defect configuration may correspond to a stable but also to a metastable state in our defect model. This aspect is pointed out in Fig. 21, where the distribution of relative stability $\varepsilon_{T1'}$ are exemplarily shown for different defects. In this figure, the atomic configurations are implicitly mapped to the

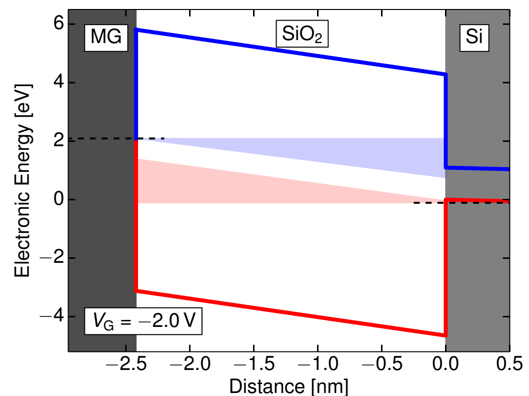


FIG. 20. Active trapping regions of a SiO_2 pMOSFET, biased at −0.5 V and −2.0 V, for the interaction with the metal gate (blue area) and substrate (red area). For the latter, this region is confined between the substrate Fermi level of the low and the high gate bias level (solid red lines) because only the defects therein are shifted above and below the Fermi level during stimulated charge trapping, can capture and emit a hole from the substrate, and therefore respond to bias switches. Defects below this area will always remain neutral while the ones above this area keep their fixed positive charge. Analogously, there also exists a blue shaded area where electron trapping from the metal gate occurs.

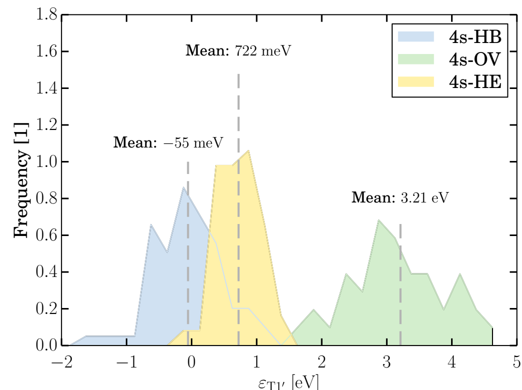


FIG. 21. Histogram of the relative stability $\varepsilon_{T1'}$ for the hydroxyl group, the hydrogen bridge, and the oxygen vacancy. Only bistable defects are considered in the statistics.

states of the four-state NMP model according to the same scheme

$$A^0 \to 1, \quad B^0 \to 1', \quad A^+ \to 2', \quad B^+ \to 2 \qquad (6.2)$$

for each defect. The used mapping scheme identifies the configuration A with the primary configuration of the four-state NMP model and consequently the configuration B must be assigned to the secondary configuration. This mapping scheme is motivated by the finding from previous DFT studies that the puckered configuration of the oxygen vacancy corresponds to the metastable state [51, 52]. These studies are also confirmed by our DFT simulations, which predict the puckered configuration B^0

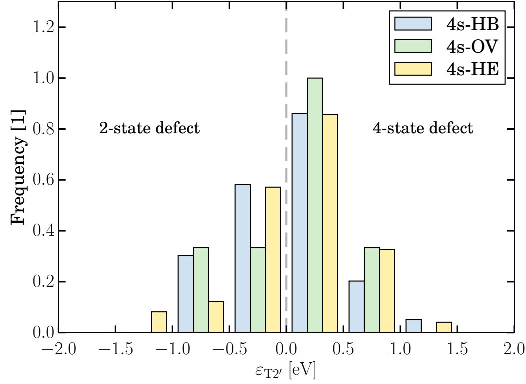


FIG. 22. Histogram of the relative stability $\varepsilon_{T2'}$ for the hydroxyl group, the hydrogen bridge, and the oxygen vacancy. The above statistics are created accounting for the relative stability between the configurations A^0 and B^0 by using either of the mapping schemes (6.2) and (6.3). It is clearly visible that $\varepsilon_{T2'}$ is distributed around zero, meaning that the state $2'$ may be higher or lower in energy. In the latter case, $\varepsilon_{T2'}$ has a negative value and the defect shows the behavior of an effective two-state trap.

Eff. 2-States	E'_t [eV]	E''_t [eV]	E_T [eV]
HB	-0.016 ± 0.399	-0.097 ± 0.451	-0.481 ± 0.449
OV	0.064 ± 0.959	-2.931 ± 0.224	-3.420 ± 0.360
HE	0.120 ± 0.390	-0.302 ± 0.347	-0.671 ± 0.418
4-States			
HB	0.266 ± 0.488	-0.453 ± 0.340	-0.104 ± 0.353
OV	0.764 ± 0.259	-2.723 ± 0.331	-2.310 ± 0.425
HE	0.760 ± 0.459	-0.310 ± 0.467	0.082 ± 0.443
Both			
HB	0.141 ± 0.472	-0.295 ± 0.431	-0.271 ± 0.440
OV	0.531 ± 0.679	-2.792 ± 0.315	-2.680 ± 0.661
HE	0.511 ± 0.534	-0.307 ± 0.424	-0.210 ± 0.568

TABLE II. Trap levels E'_t, E''_t, and E_T for the effective two-state and four-state traps of the hydrogen bridge, the oxygen vacancy, and the hydroxyl E' center. They are referenced to the valence band edge of the Si substrate and given by their mean value plus/minus one standard deviation. The values for both trap types have been given for comparison with other publications, however, it should be kept in mind that most publications do not distinguish between effective two-state and four-state defects.

of the oxygen vacancy to be more stable by 3.21 eV on average than the dimer configuration A^0. This fact is reflected in the histogram of Fig. 21, showing the positive relative stability $\varepsilon_{T1'}$ of the oxygen vacancy for the above mapping scheme. The same also holds for the hydroxyl group E' center. Its energy difference is reduced to 722 meV and thus makes the puckered configuration accessible under device operation conditions. The value of $\varepsilon_{T1'}$ is relevant for the hole emission, where a positive value triggers a transition over the metastable state $1'$ instead of state $2'$. For this transition pathway, hole emission is dominated by the gate bias dependent NMP transition $2 \leftrightarrow 1'$ and shows a switching trap behavior similar to the switching trap (see Fig. 1). In the case of the hydrogen bridge, however, $\varepsilon_{T1'}$ can assume positive and negative values. Then the above mapping scheme is inconsistent with the four-state NMP model for defects having metastable configurations A^0. In this case, the configuration A^0 cannot correspond to the initial or the final state for charge capture and emission because the configuration B^0 is lower in energy. Consequently, the defects do dwell most of their time in configurations B^0, which become the initial or final states. Then, the roles of the configurations A^0 and B^0 are actually exchanged, leading to the following mapping scheme:

$$A^0 \rightarrow 1', \quad B^0 \rightarrow 1, \quad A^+ \rightarrow 2, \quad B^+ \rightarrow 2' \tag{6.3}$$

If the correct mapping scheme (6.2) or (6.3) is selected, not only the neutral state but also the assumed positive states must be re-ordered. As a result, the choice of the mapping scheme also affects the relative stability $\varepsilon_{T2'}$ (see Fig. 22), which is used for the classification of the defects into four-state and effective two-state traps. If $\varepsilon_{T2'}$ has a positive value, the state 2 is the stable configuration and corresponds to the initial or the final state during a charge emission and capture event, respectively. Then, the trapping dynamics are dominated by transitions between the states 1 and 2 as implicitly assumed for the trap in the four-state NMP model. However, $\varepsilon_{T2'}$ may also have a negative sign, meaning that the state $2'$ actually represents the actual stable configuration. In this case, the trapping dynamics are governed by direct transitions between the stable states 1 and $2'$. Indirect transitions may take place as three-state processes involving the states $1'$ and 2 but are likely to be suppressed. This is due to the fact that either the thermal barriers $\varepsilon_{11'}, \varepsilon_{2'2}$ or the NMP barriers $\Delta V^{\ddagger}_{1'2}, \Delta V^{\ddagger}_{21'}$ are usually much larger than the NMP barriers $\Delta V^{\ddagger}_{12'}, \Delta V^{\ddagger}_{2'1}$. Therefore, these defects show an effective two-state trap behavior despite their bistability. Such traps were also encountered in our DFT simulations with a considerable frequency and have their parameter $\varepsilon_{T2'}$ widely distributed around zero with contributions from negative values, see Fig. 22 and Table III. Our data show that the majority of bistable hydrogen bridges (56%), hydroxyl E' centers (61%), and oxygen vacancies (66.7%) acts as four-state traps (see Fig. 23). However, these numbers also imply that two-state variants exist in appreciable concentrations. The existence of two-state defects seems to be consistent with a large number of publications in which the behavior of four-state as well as two-state traps was observed experimentally [9, 129].

The energy levels of both trap types are evaluated using the definitions (3.112) - (3.114). The obtained values were referenced to the substrate valence band edge, which is the source or sink of exchanged charge carriers in pMOSFETs and therefore meaningful for the inter-

Dataset	$\varepsilon_{\mathrm{T1'}}$ [eV]	$\varepsilon_{\mathrm{T2'}}$ [eV]
4s-HB	-0.055 ± 0.529	0.025 ± 0.452
4s-OV	3.210 ± 0.784	0.113 ± 0.494
4s-HE	0.722 ± 0.297	0.096 ± 0.489

TABLE III. Relative stabilities $\varepsilon_{\mathrm{T1'}}$ and $\varepsilon_{\mathrm{T2'}}$. Data are given as mean value plus/minus one standard deviation.

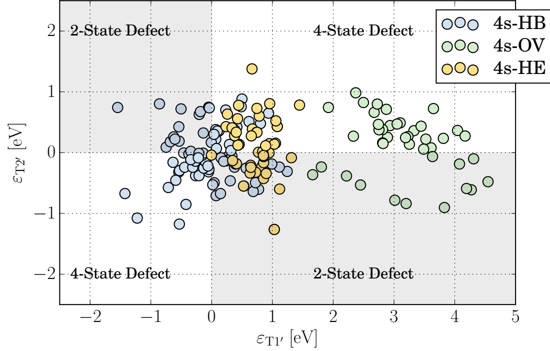


FIG. 23. Scatter plot $\varepsilon_{\mathrm{T1'}}$ [eV] vs $\varepsilon_{\mathrm{T2'}}$ [eV]. This figure demonstrates that all three defects, i.e. HB, OV, and HE, occur as effective two-state or four-state defect.

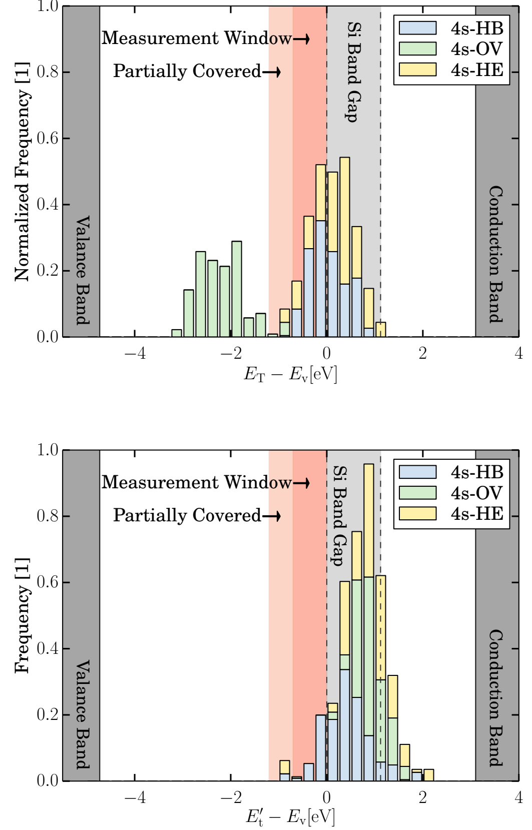


FIG. 24. The distributed energy levels E_{T} (top) and E'_{t} (bottom) of the four-state traps. The rose shaded area marks the energy region in which defects can be charged and discharged for 'typical' bias conditions during stimulated charge trapping and therefore also contribute to it. For these bias conditions, the trap level can be shifted by about 1 eV at maximum, assuming that the defect is located in the middle of a 2 nm-thick device.

pretation of charge trapping. The Si valence band offset between Si and $a{-}\mathrm{SiO_2}$ was set to 4.728 eV, which is within the range of values given in the literature [158]. It is furthermore noted that the calculation of energy levels in DFT suffers from uncertainties regarding the energetical alignment. Even though this issue has been improved by the use of hybrid functionals in our study, the energy alignment may still contain errors of up to ± 0.7 eV [132] and quantitative interpretations should be treated with caution.

In TDDS studies, four-state traps are prevalently observed and contribute to stimulated charge trapping. For $a{-}\mathrm{SiO_2}$, their trap level E_t is widely distributed in energy and shown in a histogram for our selection of defect candidates (see Fig. 24). The trap levels of the hydrogen bridge and the hydroxyl E' center are centered close to the Si substrate valence band edge (see also Tab. II). Given the large width of the distributions, a considerable fraction of the defects feature a trap level within the 'measurement window', which is relevant for 'typical' operation conditions of transistors. By contrast, the oxygen vacancy has a trap level which lies approximately -2.4 eV below the substrate valence band edge. This energy level cannot be shifted above the substrate Fermi level under realistic bias conditions for the thin oxides used today and therefore must be discarded as a possible defect candidate.

Even though positions of defect levels have been published for several gate oxides [7, 44, 50, 53, 56–59, 130, 159–162], no published values could be found for the trap level E_t of bistable defects. This is due to the fact that the trap level E_t is only defined for defects that are bistable in two charge states, and this kind of defect has only been scarcely investigated so far. However, according to equation (3.93) this trap level is strongly related to the quantity E''_{t}, which is associated with the direct transition between the states 1 and 2′ in the primary configuration. For instance, this energy level was calculated for the hydrogen bridge in crystalline $\mathrm{SiO_2}$ by Bloechl *et al.* [7]. The value of E''_{t} in this study lies approximately 0.4 eV above the valence band edge and therefore falls within the distribution of trap levels E''_{t} (-0.3 ± 0.9 eV) obtained in this work (cf. Tab. II). For the oxygen vacancy, the single-electron levels is located 2.0 eV above the $\mathrm{SiO_2}$ valence band edge in [55, 130]. This is in agreement with the corresponding value of -2.792 eV in our study, where the trap level has been measured from the Si valence band edge and equals -1.94 eV when referenced to $\mathrm{E_V}$ of $\mathrm{SiO_2}$.

The RTN at the high or the low bias level of stimulated

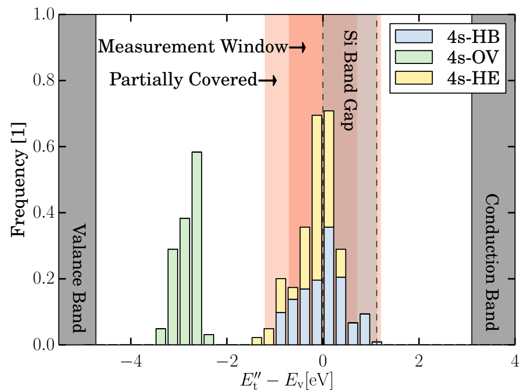


FIG. 25. The distributed energy levels E''_{t} of the two-state traps. Such traps are typically observed in noise experiments but may also play a role in stimulated charge trapping. They become RTN active if their trap level coincides with the Fermi level. The rose shaded measurement window covers the energy region relevant for 'typical' bias conditions during device operation.

charge trapping is produced by the transitions $1' \leftrightarrow 2$, which are associated with the trap level E'_{t}. This energy level must be located above the substrate Fermi level so that the defect remains positively charged when a bias is applied to the gate. However, this defect level must not be far away from the substrate Fermi level for two reasons: It should be capable of (1) being discharged after the removal of the gate bias [34] and (2) producing an RTN signal for normal low-level bias conditions. Furthermore, the trap level E'_{t} is also associated with the switching trap behavior because hole emission involves the gate bias dependent NMP transitions $1' \leftrightarrow 2$ for $E'_{\mathrm{t}} \gtrsim E_{\mathrm{f}}$. If the trap level E'_{t} is located much higher than E_{f}, the defect will essentially behave as a fixed oxide trap. The distribution of the trap level E'_{t} is depicted for our selection of defects in Fig. 24. For the oxygen vacancy, the mean value is located 0.531 eV above the substrate valence band. This result is in good agreement with the single-electron level obtained for the puckered oxygen vacancies in [51]. Bloechl *et al.* [7] obtained a thermodynamic trap level E'_{t} in crystalline $\mathrm{SiO_2}$, located 0.2 eV above midgap, and thus also compares well with our mean value of 0.531 eV. It is noted here that the trap levels E''_{t} and E'_{t} correspond to the primary and the secondary configuration of the four-state defect model. These configurations are sometimes discussed as separate defects in the literature as it was done in [130] in the case of the OV.

Besides the aforementioned four-state traps, also effective two-state traps have been observed in numerous studies on drain current RTN. This noise may be ascribed to the direct transitions $1 \leftrightarrow 2'$, linked to the thermodynamic trap level E''_{t}. As shown in Fig. 25, the oxygen vacancy has energy levels distributed around 2.931 eV below the substrate valence band with a standard deviation of 0.224 eV. Even considering the width of the distribution, none of the oxygen vacancies has a trap level which can be shifted close to the Si bandgap. As such, the oxygen vacancy must also be ruled out as a trap causing drain current RTN. By contrast, the hydrogen bridge and the hydroxyl E' center show a distribution of trap levels E''_{t} with a considerable fraction falling into the measurement window for RTN.

In lieu of the direct transition $1 \leftrightarrow 2'$, the effective two-state defects may also undergo transitions between the states 1 and $2'$ via the intermediate states $1'$ and 2. This transition pathway involves the NMP transitions $1' \leftrightarrow 2$ with the trap level E'_{t}. In this case, the noise signal is only produced if the energy minima of the initial and the final states (1 and $2'$) coincide and therefore does not depend on the position of the trap level E'_{t}. Therefore, the trap level E''_{t} remains the crucial trap level for this type of defects.

The trap level E''_{t} may be also relevant for effective two-state traps with respect to stimulated charge trapping. However, they have not been experimentally assessed by TDDS so far due to their short emission times. In addition, also four-state traps can show a similar behavior to the effective two-state traps provided that the thermal transition $2' \rightarrow 2$ proceed at larger timescales than the NMP transition $1 \leftrightarrow 2'$.

VI.2. Impact of Quantum Effects on the Temperature Dependence

Charge trapping in transistors has been demonstrated to be thermally activated by a large number of experimental investigations [25, 163, 164]. The observed thermal behavior can be traced back to a transition over an energy barrier in the configuration coordinate diagram. In the four-state NMP model, this transition can be either a pure thermal transitions ($1 \leftrightarrow 1'$ or $2 \leftrightarrow 2'$) or an NMP process ($1 \leftrightarrow 2'$ or $1' \leftrightarrow 2$). While the former can be described by simple transition state theory [40] following equation (3.102), the latter is related to the more complicated and bias-dependent lineshape function (3.51). In the classical limit, however, this complicated quantity simplifies either to equation (3.71) or to equation (3.67), depending on whether the involved parabolic potentials have the same curvature or not. Both of these expressions are governed by an exponential term including the energy barrier from the energy minimum of the parabolic potential $V_i(Q)$ up to the intersection point with the parabolic potential $V_j(Q)$. According to equations (3.59) and (3.60) for harmonic oscillators, this barrier strongly depends on the shapes of the involved potentials and is a function of their curvatures c_i and c_j and the displacement ΔQ_{ij}. In the quantum mechanical picture, these model parameters are related to the vibrational frequencies (3.61) and (3.62) or the Huang-Rhys factors (3.43) and (3.44). The latter can be directly extracted from the absorption and emission spectra of molecules and therefore have become an established quantity to describe the shape of the harmonic potentials. Since these

parameters can also be determined from our DFT data, their distributions will be studied in the following.

In the classical picture, the barrier (3.68) seems to be dependent of the displacement ΔQ_{ij}. However, this is due to a unfortunate mathematical representation of the underlying physical problem as will be demonstrated in the following: It is assumed that the displacement has two different values, ΔQ_{ij} and $\Delta\tilde{Q}_{ij}$, which differ by the factor ϑ.

$$\Delta\tilde{Q}_{ij} = \vartheta \Delta Q_{ij} \tag{6.4}$$

By contrast, the reorganization energies do not depend on the magnitude of ΔQ_{ij} since they are defined by the equations

$$S_i\hbar\omega_i = V_i(Q_j^0) - V_i(Q_i^0) \tag{6.5}$$

$$S_j\hbar\omega_j = V_j(Q_i^0) - V_j(Q_j^0) \tag{6.6}$$

as illustrated in Fig. 7. Therefore, their expressions (3.97)-(3.100) can be used to relate the curvatures $c_{i/j}$ and $\tilde{c}_{i/j}$. Using the equations (3.101) and (3.61)-(3.62), one obtains

$$\tilde{c}_i = \vartheta^2 c_i \tag{6.7}$$

$$\tilde{c}_j = \vartheta^2 c_j \ . \tag{6.8}$$

and the quantity

$$\tilde{R}_i = \frac{\tilde{c}_i}{\tilde{c}_j} \tag{6.9}$$

for the stretched displacement. As a result, the energy barrier (3.73) is not affected by the stretching factor ϑ and thus the length of the displacement ΔQ_{ij}. This implies that the classical formulation is over-specified by the combination of the three quantities ΔQ_{ij}, c_i, and c_j. Instead, the classical formulation is already determined by the reorganization energies $S_i\hbar\omega_i$ and $S_j\hbar\omega_j$ [34], which are therefore better suited to characterize the potentials $V_i(Q_i)$ and $V_j(Q_j)$. However, the potential $V_j(Q_j)$ can be discussed using the quanity R_i [65, 165], defined by

$$S_i\hbar\omega_i = R_i^2 S_j\hbar\omega_j \ , \tag{6.10}$$

in which the states i and j refer to the neutral and the positive charge state, respectively. R_i relates the reorganization energies of the involved charge states. As the latter are associated with the curvatures of the corresponding diabatic potentials, R_i can be interpreted as a measure for the deviation in the curvatures of the two charge states. For the special case $R_i = 1$, the reorganization energies ($S_i\hbar\omega_i = S_j\hbar\omega_j$) are equal, implying that the involved potentials $V_i(Q_i)$ and $V_j(Q_j)$ have the same curvatures ($c_i = c_j$). As such, the assumption of $R_i = 1$ is associated with linear electron-phonon coupling.

The distributions of the reorganization energies have been extracted from DFT simulations for our three defect candidates, as given in Tab. IV and shown in the histograms of Fig. 26 and Fig. 27. The mean reorganization energies $S_i\hbar\omega_i$ of the different defect structures

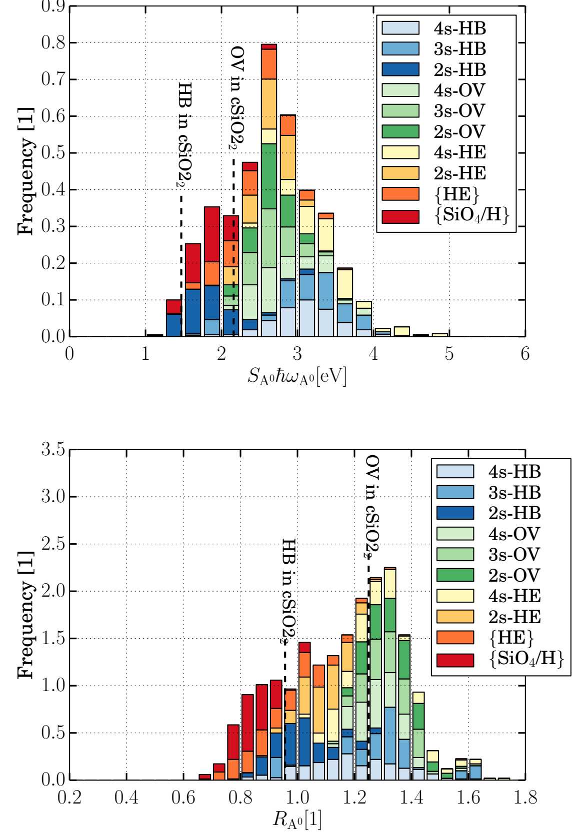


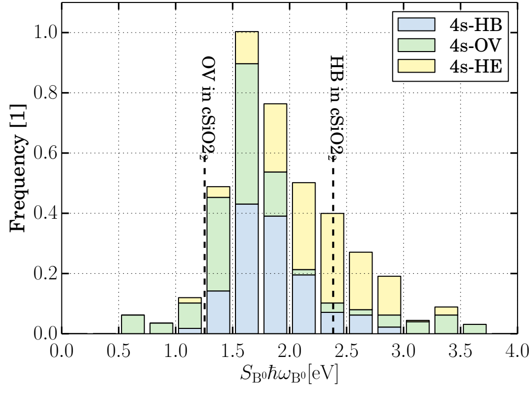
FIG. 26. Histogram of the reorganization energy $S_{A^0}\hbar\omega_{A^0}$ (top) and the model parameter R_{A^0} (bottom) in our DFT datasets.

range between 1.72 eV and 3.26 eV and are thus widely spread over 1.5 eV. However, their standard deviations remain below 0.66 eV for the investigated defect structures. Only the hydrogen bridge with a standard deviation of nearly 1 eV is an exception here. These distribution widths may appear wide, however, they are actually consistent with previous DFT calculations regarding the negatively charged hydrogen bridge [56]. It is furthermore noted that the reorganization energies calculated for the hydrogen bridge and the oxygen vacancy in crystalline SiO_2 considerably deviate from their corresponding mean values in $a{-}SiO_2$ by more than one standard deviation. This underscores the necessity to collect statistical data from realistic amorphous material systems.

The distributions of $S_i\hbar\omega_i$ in Fig. 26 contain the most frequently suspected defects in the context of charge trapping. These defects substantially differ in their chemical structure, where a considerable fraction of them even feature secondary configurations associated with additional NMP transitions. Furthermore, the amorphous SiO_2 host structure causes variations in the trap properties and thus gives rise to two-, three-, as well as four-state traps. As such, the presented statistics are considered to contain enough distinct defect structures so that the combined data give representative distributions for oxide

Dataset	$S_{A^0}\hbar\omega_{A^0}$ [eV]	R_{A^0} [1]	$S_{B^0}\hbar\omega_{B^0}$ [eV]	R_{B^0} [1]
4s-HB	3.07 ± 0.98	1.18 ± 0.20	1.72 ± 0.32	1.00 ± 0.10
3s-HB	2.95 ± 0.66	1.26 ± 0.17	—	—
2s-HB	1.73 ± 0.37	0.99 ± 0.09	—	—
4s-OV	2.68 ± 0.41	1.27 ± 0.09	1.62 ± 0.65	1.04 ± 0.28
3s-OV	2.54 ± 0.22	1.29 ± 0.07	—	—
2s-OV	2.55 ± 0.31	1.30 ± 0.08	—	—
4s-HE	3.26 ± 0.50	1.29 ± 0.26	2.12 ± 0.42	0.91 ± 0.17
2s-HE	2.48 ± 0.22	1.06 ± 0.07	—	—
{HE}	2.29 ± 0.43	0.93 ± 0.15	—	—
{SiO_4/H}	1.73 ± 0.26	0.82 ± 0.07	—	—

TABLE IV. The reorganization energies $S_{A^0}\hbar\omega_{A^0}$ and $S_{B^0}\hbar\omega_{B^0}$ along with their corresponding values R_{A^0} and R_{B^0}, all of which have been extracted from our DFT datasets. The abbreviations HB, OV, and HE stand for the hydrogen bridge, the oxygen vacancy, and the hydroxyl E' center, respectively. Their prefixes '2s', '3s', and '4s' indicate the number of stable configurations, found in the corresponding datasets. The distributions are given by their average plus/minus one standard deviation.



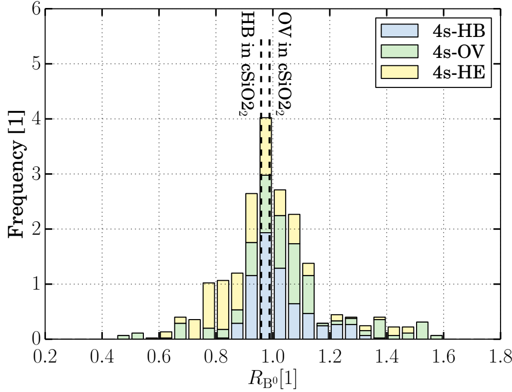


FIG. 27. The same as in Fig. 26 but for the model parameters $S_{B^0}\hbar\omega_{B^0}$ (top) and R_{B^0} (bottom).

	$S_{A^0}\hbar\omega_{A^0}$ [eV]	R_{A^0} [1]	$S_{B^0}\hbar\omega_{B^0}$ [eV]	R_{B^0} [1]
A1	3.22	1.05	0.38	0.99
A4	1.87	1.19	—	—
A6	3.32	1.14	1.48	0.94
B3	3.94	1.18	1.41	0.95
D6	1.58	0.77	—	—
G5	1.52	0.75	—	—
H4	1.78	0.77	1.30	0.71

TABLE V. Summarized model parameters $S_{A^0}\hbar\omega_{A^0}$, R_{A^0}, $S_{B^0}\hbar\omega_{B^0}$, and R_{B^0}, obtained by fitting the four-state NMP to the experimental TDDS data.The labels 'A1', 'A4', etc. refer to single-defect data obtained from TDDS studies [65] and include our showcase examples 'A1' and 'A4'.

defects in general. This argument is also corroborated by the fact that the combined distributions are found to largely overlap. For these reasons, the combined distributions give indications for realistic ranges of the model parameters $S_i\hbar\omega_i$. It has appreciable contributions varying between 1.0 eV and 4.0 eV, with an average at 2.31 eV and a standard deviation of 0.69 eV.

The mean values of the model parameter R_i lie between 0.82 and 1.30 and their corresponding standard deviations are about 0.26. As before, the values obtained for crystalline SiO_2 are found to strongly deviate from their corresponding mean values in an amorphous host material. Furthermore, the combined distributions of Fig. 26 and Fig. 27 range between 0.5 and 1.5, where the average lies at 1.06 eV and the standard deviation is 0.22 eV. This means that the curvature of the neutral potential tends to be stronger compared to that of the positive charge state.

The reorganization energy $S_i\hbar\omega_i$ and R_i are critical parameters of the four-state NMP model since they determine the temperature and the gate bias dependence of the four-state NMP model. For instance, these parameters were calibrated to the experimental data of TDDS measurements and the obtained values are listed in Tab. V. The calibrated experimental reorganization energies $S_{A^0}\hbar\omega_{A^0}$ vary between 1.52 eV and 3.94 eV and are thus well within the ranges given by our DFT simulations. Also, the parameter R_i is predicted to fall within the

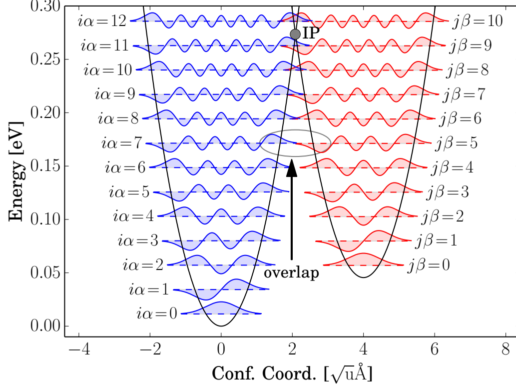


FIG. 28. The vibrational wavefunctions of two harmonic oscillators and their overlaps. The equilibrium configurations Q_i^0 and Q_j^0 are separated by $4\sqrt{\mathrm{u}}$Å and the reorganization energies $S_i\hbar\omega_i$ and $S_j\hbar\omega_j$ are set to 1 eV. The vibrational wavefunctions are plotted for the range in which the probability of finding the defect system exceeds 99.99%. The wavefunctions are found to overlap significantly at energies above 0.17 eV, with the lowest one corresponding to the transition $7 \rightarrow 5$.

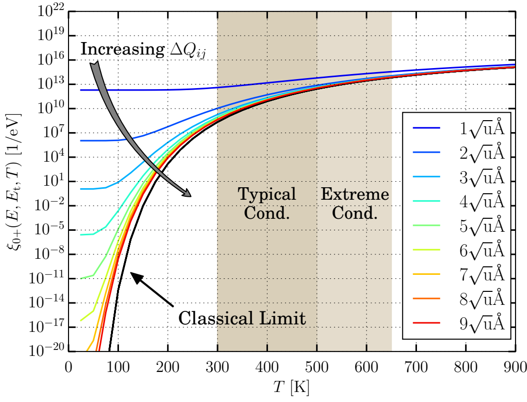


FIG. 29. Lineshape function for different displacements ΔQ_{ij}. Its trap level E_t coincides with the energy E of a charge carrier. In the band edge approximation, this quantity is proportional to the transition rate of an NMP process. In the above simulations, the reorganization energies are chosen to be 2.5 eV, which falls well into the range of values extracted from DFT simulations. It is noted that the simulated rates approach the classical limit with increasing displacement ΔQ_{ij}.

range of [0.5eV, 1.5eV]. As such, our model calibrations to TDDS data are consistent with the DFT results.

VI.2.1. High-Temperature Limit

In the classical high-temperature limit, the NMP transition is assumed to occur at the intersection point (IP) between the initial and the final potential. In the schematic of Fig. 28, the intersection lies at an energy

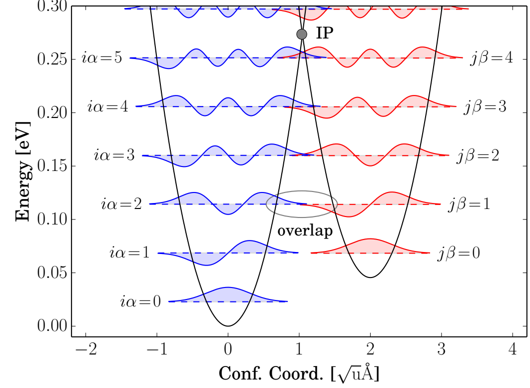


FIG. 30. The same as in Fig. 28 but for a reduced displacement of $2\sqrt{\mathrm{u}}$Å. Following equations (3.43) and (3.44), the small displacement results in a less dense energy spectrum of the harmonic oscillators but also in larger wavefunction overlaps. The energetically lowest transition has decreased to an energy of 0.11 eV, thereby reducing the effective energy barrier.

of about 0.27 eV, independently of the magnitude of the displacement ΔQ_{ij}. The quantum mechanical formulation of the CT reactions is based on harmonic oscillators, which feature a characteristic discrete energy spectrum with a constant energy spacing of $\hbar\omega_i$ and $\hbar\omega_j$. Due to the nuclear tunneling, the NMP transition can also take place below the intersection. In Fig. 28, this is the case for the transition $7 \rightarrow 5$ at an energy of 0.17 eV. As a result, the defect can undergo an NMP transition below the intersection and therefore must only overcome a reduced thermal barrier. This effect has been illustrated for the showcase example in Fig. 29, where the magnitude of the lineshape function has been plotted as a function of the temperature for a fixed pair of diabatic potentials. It is obvious that the classical approximation may lead to a severe error of the lineshape function, associated with a considerable reduction of the NMP transition probability. The error due to the classical approximation diminishes for higher temperatures but still underestimates the classical solution by about two orders of magnitude even at 300 K. This indicates that the effect of nuclear tunneling cannot be neglected for normal use-conditions of microelectronic transistors.

Notably, the effect of nuclear tunneling is also observed to become more pronounced for lower displacements ΔQ_{ij}. This is a result of the fact that the overlaps between the initial and final vibrational wavefunctions already become significant at lower energies as demonstrated by the comparison of Fig. 28 and Fig. 30. There, the displacement ΔQ_{ij} is reduced from $4\sqrt{\mathrm{u}}$Å (in Fig. 28) to $2\sqrt{\mathrm{u}}$Å, resulting in contributions from transitions $2 \rightarrow 1$ at an energy of 0.11 eV. The nuclear tunneling effect is also shown in Fig. 31 where the effective transition barrier was extracted for different values of the displacement ΔQ_{ij}. The data show that ΔQ_{ij} can significantly affect the effective barrier and may lead to tremendous

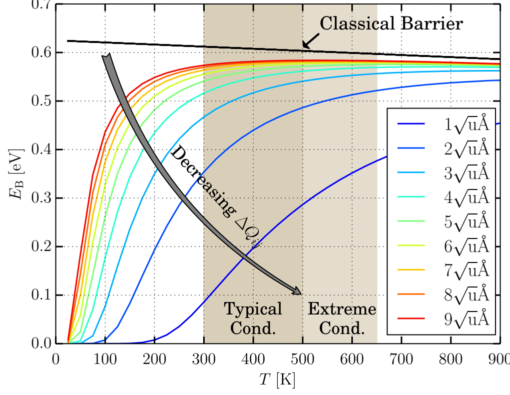


FIG. 31. Effective energy barrier calculated by using an Arrhenius law for different displacements ΔQ_{ij}. The above data are extracted from the simulations in Fig. 29. They demonstrate that the classical approximation strongly overestimates the effective energy barrier for a temperature of 400 K, which lies in the middle of the range of the 'typical' conditions.

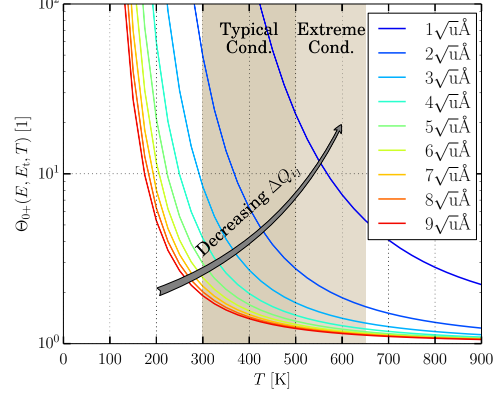


FIG. 33. Error made by neglecting the nuclear tunneling ($E_t = E_v$). If the displacement ΔQ_{ij} is larger than $3\sqrt{\mathrm{u}}$Å, the absolute error in the NMP transition rates is estimated to remain below one order of magnitude above room temperature. This error is considered significant since the capture and emission rates just vary by a few orders of magnitude within the typical conditions.

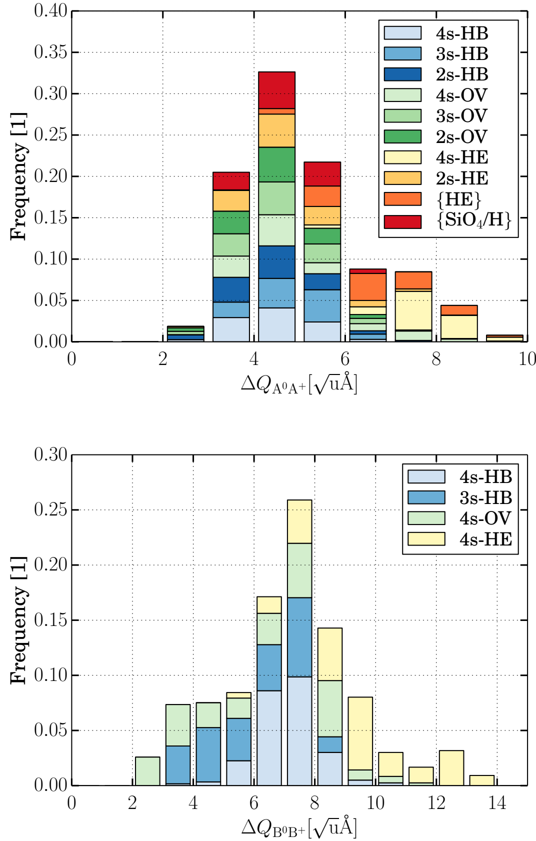


FIG. 32. The displacement ΔQ_{ij} for the configurations A and B extracted from our DFT dataset.

errors at low temperatures. Even though the tunneling effect decreases with higher temperatures, it is still pronounced in the middle range of the 'typical' conditions (around 400 K). As an example, the effective barrier is overestimated by 0.1 eV assuming a temperature of 400 K and a displacement ΔQ_{ij} of $3\sqrt{\mathrm{u}}$Å. The latter value lies at the lower limit of the ΔQ_{ij} distribution, which has its largest contributions in the range between $3\sqrt{\mathrm{u}}$Å and $10\sqrt{\mathrm{u}}$Å (cf. Fig. 32). As such, the chosen value of $3\sqrt{\mathrm{u}}$Å corresponds to the worst-case error of the classical approximation. This example shows that nuclear tunneling remains relevant for typical operation conditions of microelectronic transistors and therefore should be considered in the calculation of the NMP transition rates.

The error made by the classical approximation can also be quantified by the ratio between the quantum mechanical ($\xi_{\mathrm{qm},ij}$) and the classical ($\xi_{\mathrm{cl},ij}$) lineshape function.

$$\Theta_{ij}(E, E_t, T) = \frac{\xi_{\mathrm{qm},ij}(E, E_t, T)}{\xi_{\mathrm{cl},ij}(E, E_t, T)} \tag{6.11}$$

In Fig. 33, this quantity is plotted as a function of the temperature for different displacements ΔQ_{ij}. Its value is limited to less than one order of magnitude for temperatures higher than 300 K and a realistic displacement of $3\sqrt{\mathrm{u}}$Å. In addition, the classical error $\Theta_{ij}(E, E_t, T)$ shows a large decrease within the range of the relevant temperatures. This reflects the fact that the classical approximation leads to an appreciable deviation in the temperature dependence, which becomes more pronounced at low temperatures. In conclusion, this confirms that nuclear tunneling affects the temperature dependence as well as the magnitude of the NMP transition rates for typical reorganization energies, small displacements, and low temperatures.

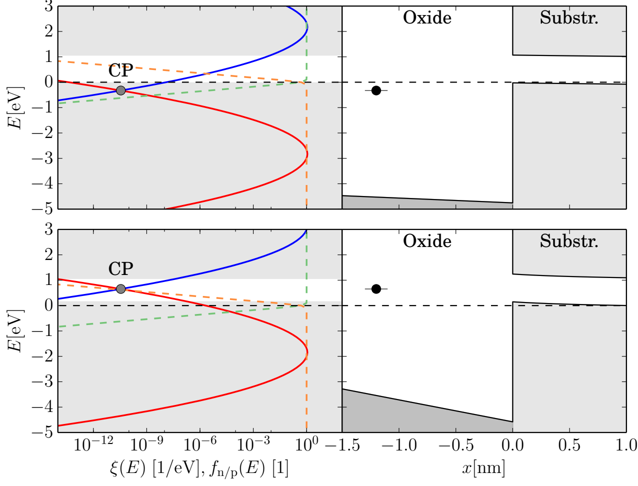


FIG. 34. Lineshape function (left) and band diagram (right) for a gate biases of $-0.4\,\mathrm{V}$ (top) and $-2.4\,\mathrm{V}$ (bottom). The electron ($f_\mathrm{n}(E)$) and the hole ($f_\mathrm{p}(E)$) occupancies are indicated by the orange and the green dashed line, respectively. As the gate bias is increased, not only the thermodynamical trap level but also the hole capture (solid red, ξ_{0+}) and hole emission (solid blue, ξ_{+0}) lineshapes are shifted upwards. The holes are energetically located close to the valence band edge $E_\mathrm{v} \approx E_\mathrm{f}$ (black dashed line). At this energy, the value of the lineshape function $\xi_{0+}(E_\mathrm{f}, E_\mathrm{t}, T)$ is increased for hole capture but reduced for hole emission when the gate bias is switched from $-2.4\,\mathrm{V}$ to $-0.4\,\mathrm{V}$. Since the integral of the NMP transition rates also has its largest contribution for strong electron-phonon coupling there, this value determines the magnitude of hole capture and emission rates. It is noted that the NMP transition remains within the strong electron-phonon coupling regime due to the large reorganization energies ($S_i\hbar\omega_i = 2.5\,\mathrm{eV}$, $S_j\hbar\omega_j = 2.5\,\mathrm{eV}$).

VI.3. Impact of Quantum Effects on the Gate Bias Dependence

In the four-state NMP model, the gate bias dependence solely originates from the NMP processes, which are described by the transition rates (3.135) and (3.136). There, the gate bias dependence enters the NMP transition rates in two ways: First, the application of a gate bias causes a shift of the Fermi level, which affects the charge carrier occupancies $f_\mathrm{n}(E)$ and $f_\mathrm{p}(E)$ within the integral of the rate equations. This effect becomes relevant during a bias sweep from accumulation into inversion or vice versa. However, transistors normally switch just between weak and strong inversion, where this effect is much less pronounced. Second, the gate bias also affects the lineshape function $\xi_{0+}(E, E_\mathrm{t}, T)$ and $\xi_{+0}(E, E_\mathrm{t}, T)$ via the energy difference between the electronic energy (E) and the thermodynamic trap levels ($E_\mathrm{t} = E'_\mathrm{t}, E''_\mathrm{t}$) as derived in Sec. III. The bias dependence of the charge carrier occupancy and the lineshape functions are illustrated in Fig. 34 for a p-channel transistor with an oxide thickness of about 2.5 nm. If the gate bias is reduced from $-0.4\,\mathrm{V}$ to $-2.4\,\mathrm{V}$, the thermodynamic trap level is raised by approximately

$$\Delta E_\mathrm{t} \approx \pm q_0 \Delta E_\mathrm{ox} x_\mathrm{t} \ , \tag{6.12}$$

where ΔE_ox is the change in the oxide field. E_t represents the trap levels of the primary (E'_t) or the secondary (E''_t) configuration. An increase of the trap level is also accompanied by an upwards energy shift of the hole capture and emission lineshape by the same amount. In particular, this can be recognized by a shift of their crossing point 'CP', which coincides with the position of the trap level. Considering hole capture, the transition probability at a certain gate bias is then given by the integral of the NMP transition rate (3.135). Its integrand is most strongly affected by the hole occupancy $f_\mathrm{p}(E)$ and the lineshape function $\xi_{0+}(E, E_\mathrm{t}, T)$. The exponent of $f_\mathrm{p}(E)$ varies linearly with the carrier energy E for thermal equilibrium between the channel charge carriers. By contrast, the exponential energy dependence of the lineshape function is much weaker since the NMP transition barriers vary weakly with ΔV_{ij}. For instance, this becomes obvious in the case of linear electron-phonon coupling (3.70), where the NMP transition barrier (3.75) becomes

$$\Delta V^\ddagger_{0+} = \left(\frac{E - E_\mathrm{t} + S_i\hbar\omega_i}{2\sqrt{S_i\hbar\omega_i}}\right)^2 \tag{6.13}$$

and shows a sublinear dependence on E in the region where $E_\mathrm{t} - E \lesssim S_i\hbar\omega_i$. As a consequence, the integral in (3.135) has its largest contribution close to the Fermi level ($E_\mathrm{f} \approx E_\mathrm{v}$). There, the occupation function saturates and the NMP hole capture rate follows the exponential dependence of the capture lineshape $\xi_{0+}(E, E_\mathrm{t}, T)$. As such, an upwards shift of the capture lineshape function results in exponentially increasing hole capture rates. Concerning hole emission, the corresponding NMP transition rate is given by equation (3.136) with the electron occupancy $f_\mathrm{n}(E)$ as the dominating factor. Analogously to hole capture, its largest contribution is found close to the Fermi level, below which the electron occupancy saturates now. The magnitude of this contribution is dominated by the exponential behavior of the emission lineshape and therefore gives rise to the decreasing emission rates with higher gate biases.

Since nuclear tunneling has been found to affect the NMP transition rates at low and medium temperatures (see the previous section), we will also address its impact on the gate bias dependence in the following. For this purpose, the lineshape function must be discussed with respect to its dependence on the displacement ΔQ_{ij}. In Fig. 35 it can be recognized that the classical lineshape function and its quantum mechanical counterpart match well in the regions around $-2.5\,\mathrm{eV}$ (regime B) and $+2.5\,\mathrm{eV}$ (regime D) while they can notably deviate in between. This region corresponds to strong electron-phonon coupling (regime C), where the vibrational wavefunctions overlap significantly below the classical intersection and therefore reduce the thermal barrier as illustrated in Fig. 34. Again, this effect becomes more pronounced for smaller displacements ΔQ_{ij}, following the

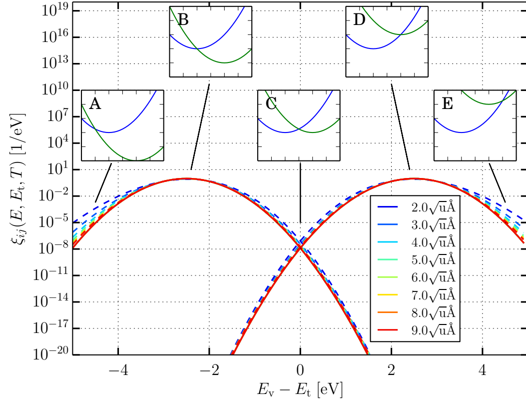


FIG. 35. Comparison between the classical (solid) and the quantum mechanical hole capture (left) and emission (right) lineshape functions ($S_i\hbar\omega_i = S_j\hbar\omega_j = 2.5\,\mathrm{eV}$, $T = 400\,\mathrm{K}$), which are plotted for different values of the displacement ΔQ_{ij}. The strong electron-phonon coupling regime covers the energy range, in which the classical intersection point lies between the minimum of the initial and the final potential as shown in the inset C. Within this regime, the classical and quantum mechanical solution differ by less than one order of magnitude. By contrast, the classical approximation leads to much larger errors for weak electron-phonon coupling, where one potential lies inside the other (see insets A and E).

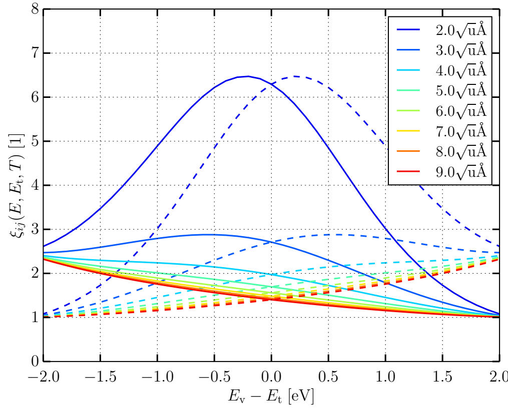


FIG. 36. Ratio between the quantum mechanical and the classical lineshape function for different displacements ΔQ_{ij}. The same defect parameters were assumed as in Fig. 35. The dashed and the solid lines correspond to the hole capture and emission lineshape function, respectively. The classical solution of the lineshape functions results in an error less than one order of magnitude.

same arguments as in the previous section. By contrast, the tunneling effect diminishes in the regions around $-2.5\,\mathrm{eV}$ (regime B) and $+2.5\,\mathrm{eV}$ (regime D) since the corresponding NMP transition barrier nearly vanishes there. The error due to the classical approximation ($\Theta_{ij}(E_\mathrm{v}, E_\mathrm{t}, T)$) is depicted in Fig. 36, which demonstrates that this approximation can lead to an underestimation of the lineshape function by a factor of up to 6.5.

With respect to the gate bias dependence, only the change in $\Theta_{ij}(E_\mathrm{v}, E_\mathrm{t}, T)$ between the upper and the lower gate bias level during stimulated charge trapping influence the accuracy of the calculated NMP transition rates. It is convenient to express this change by

$$\zeta_{ij}(E_\mathrm{t,h}, E_\mathrm{t,l}, T) = \frac{\Theta_{ij}(E_\mathrm{v}, E_\mathrm{t,h}, T)}{\Theta_{ij}(E_\mathrm{v}, E_\mathrm{t,l}, T)}\ , \qquad (6.14)$$

where $E_\mathrm{t,h}$ and $E_\mathrm{t,l}$ correspond to the trap level at the higher and the lower level of the applied gate bias, respectively. Since the quantity ζ_{ij} can have values larger or smaller than unity, the definition

$$\sigma^\mathrm{b}_{ij}(E_\mathrm{t,h}, E_\mathrm{t,l}, T) = \max_{E_\mathrm{t,h}, E_\mathrm{t,l}} \left(\zeta_{ij}(E_\mathrm{t,h}, E_\mathrm{t,l}, T), \zeta^{-1}_{ij}(E_\mathrm{t,h}, E_\mathrm{t,l}, T)\right) \qquad (6.15)$$

will be used as a measure for the classical error in the following. Under 'extreme' operation conditions, the shift of the trap level $\Delta E_\mathrm{t} = |E_\mathrm{t,h} - E_\mathrm{t,l}|$ can hypothetically reach values of up to 1 eV when the defect sits in the middle of a 2 nm-thick insulator. However, the the barriers extracted from DFT corresponding to our TDDS data are often around 0.5 eV, which is therefore assumed to correspond to 'typical' operation conditions of microelectronic transistors. As demonstrated in Fig. 36 for typical operation conditions, the ratio $\sigma^\mathrm{b}_{ij}(E_\mathrm{t,h}, E_\mathrm{t,l})$ is roughly estimated to remain below 2. Given the fact that the charge capture and emission times show an exponential dependence on the gate bias level, this factor results in a small deviation from the gate bias dependence predicted by the classical approximation. We recall that this finding has been obtained for linear electron-phonon coupling and a particular set of model parameters ($S_i\hbar\omega_i = S_j\hbar\omega_j = 2.5\,\mathrm{eV}$, $T = 400\,\mathrm{K}$) as shown in Fig. 36.

VI.4. Worst-Case Scenarios for the Classical Approximation

In the previous sections, the impact of nuclear tunneling on the temperature and gate bias dependence has been discussed using a representative defect, where the parameters $S_i\hbar\omega_i$ and R_i of this defect were set to their corresponding mean values obtained from DFT ($S_i\hbar\omega_i \sim 1.0\mathrm{eV}$, $R_i \sim 1.0$). Both the temperature and the gate bias dependence were found to be affected by quantum mechanical nuclear tunneling, albeit to a different extent. Since the dependences are expected to vary with the model parameters, worst-case scenarios for the classical approximation were investigated. For this purpose, the maximum of the classical error Θ_{ij} was determined from all combinations of temperatures T and trap level shifts ΔE_t, specified by the 'typical' or the 'extreme' operation conditions. For the latter conditions, the maximum classical error is shown for each combination of $S_i\hbar\omega_i$ and $S_j\hbar\omega_j$ in the contour plot of Fig. 37. For small relaxation energies, the magnitude of the classical

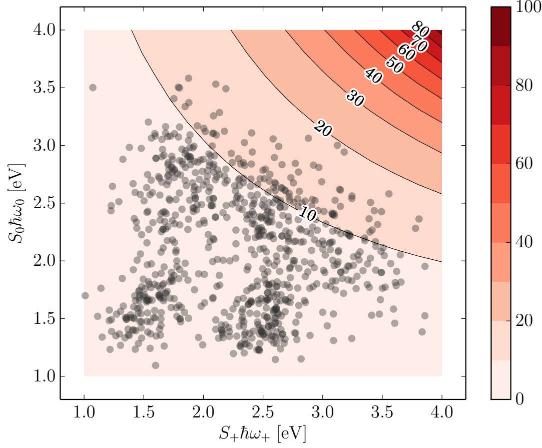


FIG. 37. The maximum classical error for different combinations of $S\hbar\omega_0$ and $S\hbar\omega_+$ for 'extreme' operation conditions. The error is represented by the shading in the contour plot while the circles represent the single defects from our DFT simulations. Even though the investigated defects cover the region with a small error for the classical approximation, this error can still yield a deviation of the hole capture rates by a factor of 20. A similar behavior is found for the hole emission rates (not shown here).

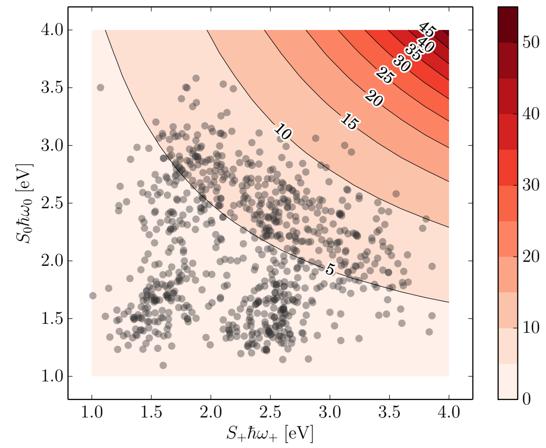


FIG. 38. The impact of the classical approximation on the temperature dependence $\sigma^{\mathrm{t}}_{ij}(E_\mathrm{t},T)$. In the above figure, 'extreme' operation conditions are investigated and the trap level is assumed to coincide with the valence band edge ($E_\mathrm{v} = E_\mathrm{t}$). Under these circumstances, the error remains below 10 for most defects (circles).

NMP transition rates must be expected to deviate from their quantum mechanical counterparts by a factor of 10. Towards larger relaxation energies, the classical error Θ_{ij} may even become more pronounced and can increase to a factor of 100. However, our collection of DFT defects accumulates in a region where the classical error remains below 10 in the worst case. The same behavior of the classical error Θ_{ij} was also observed for 'typical' operation conditions (not shown here). This can be partially ascribed to the fact that the effect of nuclear tunneling is most pronounced at low temperatures, which are covered by the 'extreme' as well as the 'typical' operation conditions.

The classical error Θ_{ij} directly enters the NMP transition rates. At a first glance, Θ_{ij} seemingly just affects the prefactor $\bar{k}$ in equations (3.135) and (3.136). In most trapping models, this prefactor is a not-further-specified quantity [35] and often interpreted as a capture or emission cross-section [8, 10, 33]. In order to investigate the impact of the classical error on the temperature dependence, the quantity

$$\sigma^{\mathrm{t}}_{ij}(E_\mathrm{t},T) = \frac{\max\limits_{T}\big(\Theta_{ij}(E_\mathrm{v},E_\mathrm{t},T)\big)}{\min\limits_{T}\big(\Theta_{ij}(E_\mathrm{v},E_\mathrm{t},T)\big)} \tag{6.16}$$

is introduced. Depending on the applied gate bias, the energy difference $|E_\mathrm{v} - E_\mathrm{t}|$ can have different values, leading to a different temperature behavior of the corresponding NMP transition rates. Therefore, the classical error in the temperature dependence must be studied for the entire energy ranges $|E_\mathrm{v} - E_\mathrm{t}| < 0.5\,\mathrm{eV}$ and $|E_\mathrm{v} - E_\mathrm{t}| < 1.0\,\mathrm{eV}$ for 'typical' and 'extreme' operation conditions, respectively. For the latter conditions, $\sigma^{\mathrm{t}}_{ij}(E_\mathrm{t},T)$ is shown in the contour plot of Fig. 38 and found to remain below a value of 10 for the large majority of the investigated DFT defects. This behavior changes only marginally for 'typical' conditions, where the error reaches a maximum value of 9. As such, the classical formulation of the NMP transition rates may result in notable errors in the predicted temperature dependence, most pronounced at room temperature.

The impact of nuclear tunneling on the gate bias dependence is investigated based on the quantity $\sigma^{\mathrm{b}}_{ij}(E_\mathrm{t,h},E_\mathrm{t,l},T)$, which is shown in the contour plot of Fig. 39 for 'typical' and 'extreme' operation conditions. For the latter conditions, the classical NMP transition rates can deviate by a factor of 4. By contrast, $\sigma^{\mathrm{b}}_{ij}(E_\mathrm{t,h},E_\mathrm{t,l},T)$ is limited to a factor of 2 for the case of 'typical' operation conditions. This deviation is small compared to the strong exponential gate bias dependence observed for the experimental hole capture and emission times. Hence, the classical formulation is found to reproduce the gate bias dependence of the NMP transition rates reasonably well.

The above findings are confirmed by our calibrations of the four-state NMP model to the experimental TDDS data of defect 'A1' (see 40). This defect shows both a pronounced gate bias and temperature dependence in its time constants and is therefore suited for investigating the effect of nuclear tunneling on the NMP transition rates. In Fig. 40, the measured capture and emission times are contrasted against a calibration of the four-state NMP model based on the quantum mechanical NMP transition rates. Due to their good agreement, the model is considered to correctly reproduce the gate bias and temperature behavior. The obtained model parameters were reused in subsequent simulations, in which the

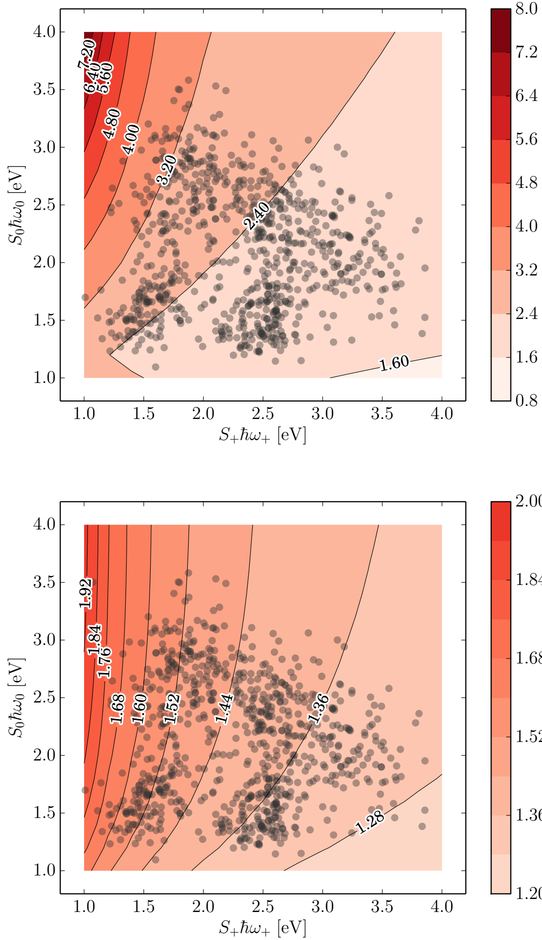


FIG. 39. The impact of the classical error on the gate bias dependence, depicted by using $\sigma_{ij}^{\mathrm{b}}(E_{\mathrm{t,h}}, E_{\mathrm{t,l}}, T)$ for 'extreme' (top) and 'typical' (bottom) conditions. Here, the temperature is set to 300 K, at which the effect of nuclear tunneling is expected to be most pronounced. The classical and the quantum mechanical formulation deviate by a factor of 4 for 'extreme' operation conditions but reduces to 2 for 'typical' operation conditions.

classical formulation of the NMP transition rates was employed and the effect of nuclear tunneling was therefore neglected. Using this parameter set, the calculated capture and emission times now fail to reproduce the experimental behavior. However, this can be traced back to the deviation in the magnitude of the NMP transition rates. If this error is compensated by a corresponding correction factor in the classical simulations, good agreement with the experimental data is achieved again. However, the temperature dependence is still underestimated, which is consistent with our findings from above. The obtained deviation may be corrected by a temperature-dependent prefactor of the classical NMP transition rates, in which the nuclear tunneling effect is estimated by a WKB approximation, similar to [94].

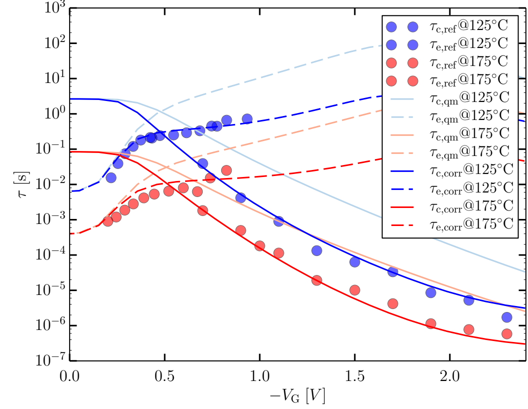


FIG. 40. Correction of the classical simulations. In a first step, the parameters were obtained by calibrating the quantum mechanical model to the experimental TDDS data of trap 'A1'. Then these parameters have been reused for a simulation with the classical formulation of the NMP transition rates. The obtained hole capture and emission times are represented by the bright lines and contrasted against the experimental data (circles) in the above figure. A multiplication of the classical rates by a correction factor (30) leads to the capture and emission times shown by the dark lines.

VII. CONCLUSIONS

In modern semiconductor devices, in particular, nanoscaled transistors, charge trapping leads to drifts and fluctuations in the drain current. Both can pose severe problems for device engineers and therefore have attracted a lot of interest over the last decades. The nature of charge trapping has recently been understood in unprecedented detail using the four-state NMP model, in which charge capture and emission is a combination thermally activated transitions and NMP processes. We have presented a rigorous description of the four-state NMP model, starting from the theory of charge trapping (CT) reaction and finally arriving at the capture and emission times that can be compared to experimental results. The NMP transitions have been described within the framework of CT reactions, the standard theory in the field of physical chemistry. The corresponding diabatic potentials have been thoroughly incorporated in the configuration coordinate diagram of the four-state NMP model, thereby relating the thermodynamic trap levels to particular points of the diabatic potentials. Interestingly, this methodology has revealed that electron as well as hole traps — both observed experimentally — can be viewed as different representations of the same defect. As such, it is conceivable that the same kind of defect is involved in electron and hole trapping in transistors.

The above methodology also allows one to extract the model parameters, such as trap levels (E_{T}, E_{t}', E_{t}'') and reorganization energies ($S_i\hbar\omega_i$, R_i), from DFT calculations. The trap levels are critical parameters in any physics-based defect model and must lie within certain energy ranges compatible with the experimentally ob-

served trapping behavior. Using their energetical position as a criterion, the oxygen vacancy — a frequently suggested defect in the context of charge trapping phenomena — must be discarded. By contrast, the hydrogen bridge as well as the hydroxyl E' center have been found to have reasonable distributions of their trap levels E_T and E'_t, confirming them as possible defects involved in charge trapping.

Furthermore, the reorganization energies have been determined from our large set of DFT simulations. The calculated range of the reorganization energies provides valuable information for calibrating the four-state NMP model to experimental data. Using our DFT calculations, the inherent uncertainties of the calibration procedure have been eliminated by determining realistic parameter ranges, which makes the prediction of the four-state NMP model more reliable.

In addition, the calculated parameter ranges have also been used to evaluate the effect of nuclear tunneling on the gate bias and temperature dependence of NMP transition rates. Our study has revealed that the activation energies are significantly underestimated by the classical approximation. However, this deviation can be well corrected by adjusting the unknown prefactor or the capture and emission cross-sections. By contrast, the classical formulation of the NMP transitions correctly reproduces the gate bias dependence for the 'typical' use-conditions of modern microelectronic transistors.

VIII. ACKNOWLEDGEMENTS

This work has received funding from the Austrian Science Fund (FWF) project n°23390-N24, the Christian Doppler Laboratory (CDL) for Reliability Issues in Microelectronics, the European Communitys FP7 n°261868 (MORDRED), the European Communitys FP7 n°619234 (MoRV), and the Intel Sponsored Research Project n°2013111914. A.L.S. acknowledges funding provided by EPSRC under grants No. EP/K01739X/1 and EP/P013503/1 and by the Leverhulme Trust RPG-2016-135. Computer facilities on Archer service have been provided via the UKs HPC Materials Chemistry Consortium (EPSRC Grant No. EP/L000202).

APPENDIX

VIII.1. Hole Trap in the Hole Picture

The description of theoretical solid-state physics naturally relies on the basic interactions between electrons and therefore the calculated energies usually refer to electrons. Along this line, our defect model has also been derived in the picture of electrons. As for semiconductors and especially p-channel transistors, the concept of holes has proven to be practical. It is based on the interactions of holes and allows for a simpler description. In this hole picture, the atomic system actually remains the same as in the case of the electron picture — but with the difference that a hole, instead of an electron, is exchanged between the defect and the substrate. Therefore, the hole resides in the defect for the positive charge state while it is located in the substrate for the neutral charge state (cf. Fig. 41).

In the hole picture, the hole energy $\bar{E}$ is defined by

$$\bar{E} = V_0(Q_1) - \bar{V}(Q_1) \tag{8.1}$$
$$= V_0(Q_{1'}) - \bar{V}(Q_{1'}) \;. \tag{8.2}$$

$\bar{V}(Q)$ denotes the diabatic potential of the atomic system, from which the positive charge is removed to infinity. Due to this definition, the electron and hole energies can be converted using the equation

$$\bar{E} = -E \;. \tag{8.3}$$

The driving forces are expressed as

$$\Delta V_{0+}^{12'} = V_+(Q_{2'}) - V_0(Q_1) \tag{8.4}$$
$$\Delta V_{0+}^{1'2} = V_+(Q_2) - V_0(Q_{1'}) \tag{8.5}$$
$$\Delta V_{0+}^{12} = V_+(Q_2) - V_0(Q_1) \tag{8.6}$$

and therefore remain unchanged in comparison to the electron picture. Inserting equation (8.1) into the above driving forces yields

$$\Delta V_{0+}^{12'} = \bar{E}''_\mathrm{t} - \bar{E} \tag{8.7}$$
$$\Delta V_{0+}^{1'2} = \bar{E}'_\mathrm{t} - \bar{E} \tag{8.8}$$
$$\Delta V_{0+}^{12} = \bar{E}_\mathrm{T} - \bar{E} \;, \tag{8.9}$$

where the corresponding trap levels are obtained as

$$\bar{E}''_\mathrm{t} = V_+(Q_{2'}) - \bar{V}(Q_1) \tag{8.10}$$
$$\bar{E}'_\mathrm{t} = V_+(Q_2) - \bar{V}(Q_{1'}) \tag{8.11}$$
$$\bar{E}_\mathrm{T} = V_+(Q_2) - \bar{V}(Q_1) \;. \tag{8.12}$$

Analogously to the electronic energies E and $\bar{E}$, the above trap levels can be converted into the electron picture using the following equations

$$\bar{E}''_\mathrm{t} = -E''_\mathrm{t} \tag{8.13}$$
$$\bar{E}'_\mathrm{t} = -E'_\mathrm{t} \tag{8.14}$$
$$\bar{E}_\mathrm{T} = -E_\mathrm{T} \;. \tag{8.15}$$

When using the definitions of the stabilities

$$\varepsilon_{\mathrm{T}2'} = V_+(Q_{2'}) - V_+(Q_2) \tag{8.16}$$
$$\varepsilon_{\mathrm{T}1'} = V_0(Q_{1'}) - V_0(Q_1) \;, \tag{8.17}$$

the following relations are found:

$$\bar{E}'_\mathrm{t} = \bar{E}_\mathrm{T} - \varepsilon_{\mathrm{T}1'} \tag{8.18}$$
$$\bar{E}''_\mathrm{t} = \bar{E}_\mathrm{T} + \varepsilon_{\mathrm{T}2'} \tag{8.19}$$

It is pointed out that each of the above expressions in the hole picture has its correspondence in the electron picture and can be converted by making use of the equations (8.3) and (8.13)-(8.15).

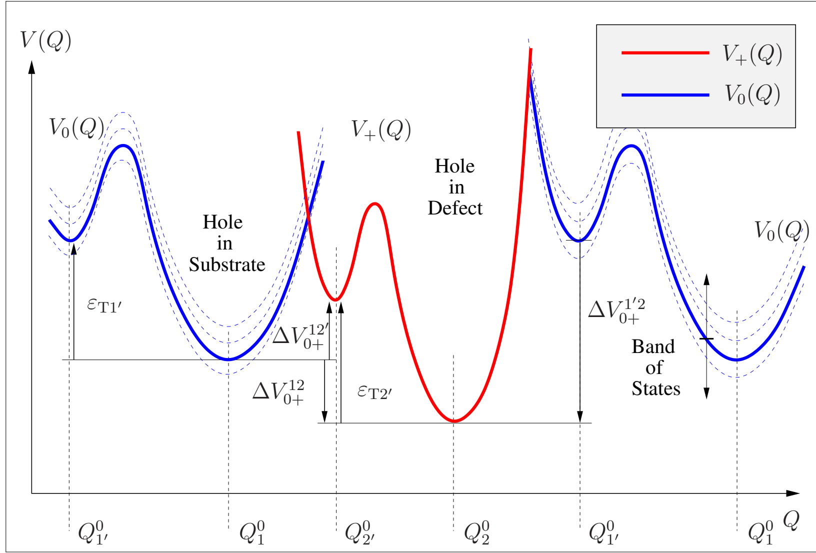


FIG. 41. Configuration coordinate diagram of a hole trap in the hole picture. In the positive charge state (red solid curve), the hole dwells within the defect but is located in the substrate when the defect is neutral (blue solid curve). Since the hole energy E may have different values within the hole picture, the vertical position of the corresponding potential $V_0(Q)$ can vary (indicated by the blue dashed curves).

VIII.2. Wenzel-Kramers-Brillouin Method

The WKB method [166] is an semiclassical approximation to compute the wavefunctions of the stationary Schrödinger equation. However, this approximation yields a diverging expression for the wavefunction, where the particle energy E equals the potential energy $V(x)$. As a result, the wavefunction left and right to this point cannot be adjusted, which is the case at the discontinuity of the semiconductor-dielectric interface for instance. One way to overcome this problem is to apply Langer's procedure [166], which is not presented here. The WKB method also applies to classically forbidden regions where the particle energy E lies below the potential barrier $V(x)$. It is therefore frequently used to approximate the wavefunctions of the substrate electrons or holes. In the classical forbidden region, the shape of the wavefunction is dominated by the exponential term

$$\lambda(x) \approx \exp\big(-\frac{1}{\hbar}\int_{x_1}^{x_2} p(x)\ \mathrm{d}x\big) \tag{8.20}$$

with

$$p(x) = \sqrt{|2m_\mathrm{t}(V(x)-E)|} \tag{8.21}$$

and m_t being the tunneling mass. x_1 and x_2 stand for the classical turning point at the semiconductor-dielectric interface and the position of the trap, respectively. Supposing that only a negligible amount of charges is located in the dielectric, the potential energy $V(x)$ can be expressed as

$$V(x) = V_1 + \underbrace{\frac{V_2-V_1}{x_2-x_1}}_{=q_0E_\mathrm{ox}}\ (x-x_1)\,. \tag{8.22}$$

For a trapezoidal barrier, $\lambda(x)$ simplifies to

$$\lambda(x) \approx \exp\big(-\kappa(V_2-E)^{\frac{3}{2}} + \kappa(V_1-E)^{\frac{3}{2}}\big) \tag{8.23}$$

with

$$\kappa = \frac{2\sqrt{2m_\mathrm{t}}}{3\hbar q_0 E_\mathrm{ox}} \tag{8.24}$$

If tunneling occurs through a triangular barrier, the classically forbidden region extends to

$$x_0 = x_1 + \frac{E-V_1}{q_0E_\mathrm{ox}}\ . \tag{8.25}$$

For negative electric fields ($V_2 < E < V_1$), one obtains

$$\lambda(x) \approx \exp\big(\kappa\ (V_1-E)^{\frac{3}{2}}\big)\ . \tag{8.26}$$

while positive electric fields ($V_1 < E < V_2$) results in

$$\lambda(x) \approx \exp\big(-\kappa\ (V_2-E)^{\frac{3}{2}}\big)\ . \tag{8.27}$$

These two cases are commonly known as the Fowler-Nordheim formulas [167]. For a rectangular barrier with $E < V_1 = V_2 = V$, the integral in equation (8.20) simplifies to a multiplication.

$$\lambda(x) \approx \exp\big(-\frac{\sqrt{2m_\mathrm{t}(V-E)}}{\hbar}(x_2-x_1)\big) \tag{8.28}$$

[1] P. Lenahan and J. Conley, "What Can Electron Paramagnetic Resonance Tell Us about the Si/SiO_2 System?," *J.Vac.Sci.Technol.B*, vol. 16, no. 4, pp. 2134–2153, 1998.

[2] P. Lenahan, "Dominating Defects in the MOS System: P_b and E' Centers," in *Defects in Microelectronic Materials and Devices* (D. Fleetwood, R. Schrimpf, and S. Pantelides, eds.), Taylor and Francis/CRC Press, 2008. (invited).

[3] J. Conley and P. Lenahan, "Room Temperature Reactions Involving Silicon Dangling Bond Centers and Molecular Hydrogen in Amorphous SiO_2 Thin Films on Silicon," *Appl.Phys.Lett.*, vol. 62, no. 1, pp. 40–42, 1993.

[4] J. Campbell and P. Lenahan, "NBTI: An Atomic-Scale Defect Persective," in *Proc.IRPS*, pp. 442–447, 2006.

[5] J. Campbell, P. Lenahan, A. Krishnan, and S. Krishnan, "Identification of the Atomic-Scale Defects Involved in the Negative Bias Temperature Instability in Plasma-Nitrided P-Channel Metal-Oxide-Silicon Field-Effect Transistors," *J.Appl.Phys.*, vol. 103, p. 044505, 2008.

[6] J. McPherson, "Quantum Mechanical Treatment of Si-O Bond Breakage in Silica Under Time Dependent Dielectric Breakdown Testing," pp. 209–216, 2007.

[7] P. Blöchl and J. Stathis, "Hydrogen Electrochemistry and Stress-Induced Leakage Current in Silica," *Phys.Rev.Lett.*, vol. 83, no. 2, pp. 372–375, 1999.

[8] L. Vandelli, A. Padovani, L. Larcher, R. Southwick, W. Knowlton, and G. Bersuker, "A Physical Model of the Temperature Dependence of the Current Through SiO_2/HfO_2 Stacks," *IEEE Trans.Elect.Dev.*, vol. 58, no. 9, pp. 2878–2887, 2011.

[9] M. Toledano-Luque, B. Kaczer, E. Simoen, R. Degraeve, J. Franco, P. Roussel, T. Grasser, and G. Groeseneken, "Correlation of Single Trapping and Detrapping Effects in Drain and Gate Currents of Nanoscaled nFETs and pFETs," in *Proc.IRPS*, 2012.

[10] M. Kirton and M. Uren, "Noise in Solid-State Microstructures: A New Perspective on Individual Defects, Interface States, and Low-Frequency (1/f) Noise," *Adv.Phys.*, vol. 38, no. 4, pp. 367–486, 1989.

[11] D. Schroder, "Negative Bias Temperature Instability: What Do We Understand?," *Microelectron.Reliab.*, vol. 47, no. 6, pp. 841–852, 2007.

[12] G. Rescher, G. Pobegen, T. Aichinger, and T. Grasser, "On the subthreshold drain current sweep hysteresis of 4h-sic nmosfets," in *2016 IEEE International Electron Devices Meeting (IEDM)*, pp. 10.8.1–10.8.4, 2016.

[13] P. Lagger, M. Reiner, D. Pogany, and C. Ostermaier, "Comprehensive study of the complex dynamics of forward bias-induced threshold voltage drifts in gan based mis-hemts by stress/recovery experiments," *IEEE Trans.Elect.Dev.*, vol. 61, no. 4, pp. 1022–1030, 2014.

[14] Y. Illarionov, G. Rzepa, M. Waltl, T. Knobloch, A. Grill, M. M. Furchi, T. Mueller, and T. Grasser, "The Role of Charge Trapping in MoS_2/SiO_2 and MoS_2/hBN Field-Effect Transistors," *2D Materials*, vol. 3, no. 3, p. 035004, 2016.

[15] Y. Illarionov, M. Waltl, G. Rzepa, J.-S. Kim, S. Kim, A. Dodabalapur, D. Akinwande, and T. Grasser, "Long-Term Stability and Reliability of Black Phosphorus Field-Effect Transistors," *ACS Nano*, vol. 10, no. 10, pp. 9543–9549, 2016.

[16] S. Zafar, A. Callegari, E. Gusev, and M. V. Fischetti, "Charge Trapping in High-K Gate Dielectric Stacks," in *Proc.IEDM*, pp. 517–520, 2002.

[17] K. Onishi, R. Choi, C. S. Kang, H.-J. Cho, Y. H. Kim, R. E. Nieh, J. Han, S. A. Krishnan, M. S. Akbar, and J. C. Lee, "Bias-Temperature Instabilities of Polysilicon Gate HfO_2 MOSFETs," *IEEE Elect.Dev.*, vol. 50, no. 6, pp. 1517–1524, 2003.

[18] C. Shen, M. F. Li, X. P. Wang, H. Y. Yu, Y. P. Feng, A. T. L. Lim, Y. C. Yeo, D. S. H. Chan, and D. L. Kwong, "Negative U-Traps in HfO : 2 Gate Dielectrics and Frequency Dependence of Dynamic BTI in MOSFETs," in *Proc.IEDM*, pp. 733–736, 2004.

[19] T. Grasser, M. Waltl, Y. Wimmer, W. Goes, R. Kosik, G. Rzepa, H. Reisinger, G. Pobegen, A. El-Sayed, A. Shluger, and B. Kaczer, "Gate-Sided Hydrogen Release as the Origin of "Permanent" NBTI Degradation: From Single Defects to Lifetimes," in *Proc.IEDM*, 2015.

[20] T. Aichinger, M. Nelhiebel, and T. Grasser, "Unambiguous Identification of the NBTI Recovery Mechanism using Ultra-Fast Temperature Changes," in *Proc.IRPS*, pp. 2–7, 2009.

[21] T. Aichinger and M. Nelhiebel, "Advanced Energetic and Lateral Sensitive Charge Pumping Profiling Methods for MOSFET Device Characterization — Analytical Discussion and Case Studies," *IEEE Trans.Dev.Mater.Rel.*, vol. 8, no. 3, pp. 509–518, 2008.

[22] A. Kerber, K. Maitra, A. Majumdar, M. Hargrove, R. J. Carter, and E. A. Cartier, "Characterization of Fast Relaxation During BTI Stress in Conventional and Advanced CMOS Devices with HfO_2/TiN Gate Stacks," *IEEE Elect.Dev.*, vol. 55, no. 11, pp. 3175–3183, 2008.

[23] H. Reisinger, U. Brunner, W. Heinrigs, W. Gustin, and C. Schlunder, "A Comparison of Fast Methods for Measuring NBTI Degradation," *IEEE Trans.Dev.Mater.Rel.*, vol. 7, no. 4, pp. 531–539, 2007.

[24] B. Kaczer, T. Grasser, P. Roussel, J. Martin-Martinez, R. O'Connor, B. O'Sullivan, and G. Groeseneken, "Ubiquitous Relaxation in BTI Stressing – New Evaluation and Insights," in *Proc.IRPS*, pp. 20–27, 2008.

[25] M. Waltl, A. Grill, G. Rzepa, W. Goes, J. Franco, B. Kaczer, J. Mitard, and T. Grasser, "Superior NBTI in High-K SiGe Transistors — Part I: Experimental," *IEEE Elect.Dev.*, vol. 64, pp. 2092–2098, 2017.

[26] M. Denais, C. Parthasarathy, G. Ribes, Y. Rey-Tauriac, N. Revil, A. Bravaix, V. Huard, and F. Perrier, "On-the-Fly Characterization of NBTI in Ultra-Thin Gate Oxide pMOSFETs," in *Proc.IEDM*, pp. 109–112, 2004.

[27] E. N. Kumar, V. D. Maheta, S. Purawat, A. E. Islam, C. Olsen, K. Ahmed, M. A. Alam, and S. Mahapatra, "Material Dependence of NBTI Physical Mechanism in Silicon Oxynitride (SiON) p-MOSFETs: A Comprehensive Study by Ultra-Fast On-The-Fly (UF-OTF) IDLIN Technique," in *Proc.IEDM*, pp. 809–812, 2007.

[28] T. Yang, C. Shen, M.-F. Li, C. Ang, C. Zhu, Y.-C. Yeo, G. Samudra, S. Rustagi, M. Yu, and D.-L. Kwong, "Fast DNBTI Components in p-MOSFET with SiON Dielectric," *IEEE Elect.Dev.Let.*, vol. 26, no. 11, pp. 826–828, 2005.

[29] I. Lundstrom and C. Svensson, "Tunneling to Traps in Insulators," *J.Appl.Phys.*, vol. 43, no. 12, pp. 5045–

5047, 1972.

[30] T. Tewksbury, *Relaxation Effects in MOS Devices due to Tunnel Exchange with Near-Interface Oxide Traps.* Ph.D. Thesis, MIT, 1992.

[31] A. McWhorter, "1/f Noise and Germanium Surface Properties," in *Sem.Surf.Phys*, RH Kingston (Univ Penn Press), 1957.

[32] S. Christensson, I. Lundström, and C. Svensson, "Low Frequency Noise in MOS Transistors — I Theory," *Sol.-St.Electr.*, vol. 11, pp. 797–812, 1968.

[33] N. Zanolla, D. Siprak, P. Baumgartner, E. Sangiorgi, and C. Fiegna, "Measurement and Simulation of Gate Voltage Dependence of RTS Emission and Capture Time Constants in MOSFETs," in *Ultimate Integration of Silicon*, pp. 137–140, 2008.

[34] T. Grasser, "Stochastic Charge Trapping in Oxides: From Random Telegraph Noise to Bias Temperature Instabilities," *Microelectron.Reliab.*, vol. 52, no. 1, pp. 39–70, 2012.

[35] A. Avellan, D. Schroeder, and W. Krautschneider, "Modeling Random Telegraph Signals in the Gate Current of Metal-Oxide-Semiconductor Field Effect Transistors after Oxide Breakdown," *J.Appl.Phys.*, vol. 94, no. 1, pp. 703–708, 2003.

[36] A. Palma, A. Godoy, J. A. Jimenez-Tejada, J. E. Carceller, and J. A. Lopez-Villanueva, "Quantum Two-Dimensional Calculation of Time Constants of Random Telegraph Signals in Metal-Oxide-Semiconductor Structures," *Phys.Rev.B*, vol. 56, no. 15, pp. 9565–9574, 1997.

[37] S. Ganichev, W. Prettl, and I. Yassievich, "Deep Impurity-Center Ionization by Far-Infrared Radiation," *Phys.Solid State*, vol. 39, no. 1, pp. 1703–1726, 1997.

[38] S. Makram-Ebeid and M. Lannoo, "Quantum Model for Phonon-Assisted Tunnel Ionization of Deep Levels in a Semiconductor," *Phys.Rev.B*, vol. 25, no. 10, pp. 6406–6424, 1982.

[39] P. Barbara, T. Meyer, and M. Ratner, "Contemporary Issues in Electron Transfer Research," *Journal of Physical*, vol. 100, pp. 13148–13168, 1996.

[40] A. Nitzan, *Chemical Dynamics in Condensed Phases: Relaxation, Transfer, and Reactions in Condensed Molecular Systems.* Oxford University Press, 2006.

[41] V. May and O. Kuhn, *Charge and Energy Transfer Dynamics in Molecular Systems.* Wiley-VCH, 2011. ISBN 978-3-527-40732-3.

[42] T. Grasser, M.Waltl, W. Goes, Y. Wimmer, A.-M. El-Sayed, A. Shluger, and B. Kaczer, "On the Volatility of Oxide Defects: Activation, Deactivation, and Transformation," in *Proc.IRPS*, pp. 5A.3.1–5A.3.8, 2015.

[43] S. Guo, R. Wang, D. Mao, Y. Wang, and R. Huang, "Anomalous Random Telegraph Noise in Nanoscale Transistors as Direct Evidence of Two Metastable States of Oxide Traps," *Scientific Reports*, vol. 7, no. 11, 2017.

[44] J. Robertson, "High Dielectric Constant Gate Oxides for Metal Oxide Si Transistors," *Rep.Prog.Phys.*, vol. 69, pp. 327–396, 2006.

[45] A. Stirling, A. Pasquarello, and J.-C. J.-C. C. Car, "Dangling Bond Defects at Si-SiO_2 Interfaces: Atomic Structure of the P_{b1} Center," *Phys.Rev.Lett.*, vol. 85, no. 13, pp. 2773–2776, 2000.

[46] S. N. Rashkeev, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, "Defect Generation by Hydrogen at the $Si-SiO_2$ Interface," *Phys.Rev.Lett.*, vol. 87, no. 16, p. 165506, 2001.

[47] L. Tsetseris and S. Pantelides, "Migration, Incorporation, and Passivation Reactions of Molecular Hydrogen at the $Si-SiO_2$ Interface," *Phys.Rev.B*, vol. 70, no. 24, p. 245320, 2004.

[48] A. Stirling and A. Pasquarello, "First-Principles Modeling of Paramagnetic Si Dangling-Bond Defects in Amorphous SiO_2," *Phys.Rev.B*, vol. 66, p. 245201, 2002.

[49] S. Mukhopadhyay, P. Sushko, A. Stoneham, and A. Shluger, "Modeling of the Structure and Properties of Oxygen Vacancies in Amorphous Silica," *Phys.Rev.B*, vol. 70, no. 19, p. 195203, 2004.

[50] S. Mukhopadhyay, P. Sushko, V. Mashkov, and A. Shluger, "Spectroscopic Features of Dimer and Dangling Bond E' Centres in Amorphous Silica," *J.Phys.-Condens.Matter*, vol. 17, p. 13111318, 2005.

[51] Z.-Y. Lu, C. Nicklaw, D. Fleetwood, R. Schrimpf, and S. Pantelides, "Structure, Properties, and Dynamics of Oxygen Vacancies in Amorphous SiO_2," *Phys.Rev.Lett.*, vol. 89, no. 28, p. 285505, 2002.

[52] M. Boero, A. Pasquarello, J. Sarnthein, and R. Car, "Structure and Hyperfine Parameters of E'_1 Centers in α-Quartz and in Vitreous SiO_2," *Phys.Rev.Lett.*, vol. 78, no. 5, pp. 887–890, 1997.

[53] A. Mysovsky, P. Sushko, S. Mukhopadhyay, A. Edwards, and A. Shluger, "Calibration of Embedded-Cluster Method for Defect Studies in Amorphous Silica," *Phys.Rev.B*, vol. 69, no. 8, p. 085202, 2004.

[54] S. Pantelides, Z.-Y. Lu, C. Nicklaw, T. Bakos, S. Rashkeev, D. Fleetwood, and R. Schrimpf, "The E' Center and Oxygen Vacancies in SiO_2," *J.Non-Cryst.Solids*, vol. 354, pp. 217–223, 2008.

[55] P. Sushko, S. Mukhopadhyay, A. Mysovsky, V. Sulimov, A. Taga, and A. Shluger, "Structure and Properties of Defects in Amorphous Silica: New Insights from Embedded Cluster Calculations," *J.Phys.-Condens.Matter*, vol. 17, no. 21, p. S2115, 2005.

[56] A. Alkauskas and A. Pasquarello, "Alignment of Hydrogen-Related Defect Levels at the $Si-SiO_2$ Interface," *Phys.B Condens.Matter*, vol. 401-402, pp. 546–549, 2007.

[57] A. Foster, F. L. Gejo, A. Shluger, and R. Nieminen, "Vacancy and Interstitial Defects in Hafnia," *Phys.Rev.B*, vol. 65, p. 174117, 2002.

[58] J. Godet and A. Pasquarello, "Protons at the $Si-SiO_2$ Interface: a First Principle Investigation," *Microelectron.Eng.*, vol. 84, no. 9-10, pp. 2035–2038, 2007.

[59] J. L. Gavartin, L. Fonseca, G. Bersuker, and A. Shluger, "Ab Initio Modeling of Structure and Defects at the HfO_2/Si Interface," *Microelectron.Eng.*, vol. 80, no. 1, pp. 412–415, 2005.

[60] S. Joeng and A. Oshiyama, "Atomic and Electronic Structures of N-Incorporated Si Oxides," *Phys.Rev.B*, vol. 86, no. 16, p. 3574, 2001.

[61] T. Grasser, H. Reisinger, P.-J. Wagner, and B. Kaczer, "The Time Dependent Defect Spectroscopy for the Characterization of Border Traps in Metal-Oxide-Semiconductor Transistors," *Phys.Rev.B*, vol. 82, no. 24, p. 245318, 2010.

[62] H. Reisinger, T. Grasser, W. Gustin, and C. Schlünder, "The Statistical Analysis of Individual Defects Constituting NBTI and its Implications for Modeling DC- and AC-Stress," in *Proc.IRPS*, pp. 7–15, 2010.

[63] T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P.-J. Wagner, F. Schanovsky, J. Franco, P. Roussel, and M. Nelhiebel,

"Recent Advances in Understanding the Bias Temperature Instability," in *Proc.IEDM*, pp. 82–85, 2010. (invited).

[64] T. Grasser, B. Kaczer, H. Reisinger, P.-J. Wagner, and M. Toledano-Luque, "On the Frequency Dependence of the Bias Temperature Instability," in *Proc.IRPS*, pp. XT.8.1–XT.8.7, 2012.

[65] T. Grasser, H. Reisinger, P.-J. Wagner, F. Schanovsky, W. Goes, and B. Kaczer, "The Time Dependent Defect Spectroscopy (TDDS) for the Characterization of the Bias Temperature Instability," in *Proc.IRPS*, pp. 16 – 25, 2010.

[66] M. Waltl, W.Goes, K. Rott, H. Reisinger, and T. Grasser, "A Single-Trap Study of PBTI in SiON nMOS Transistors: Similarities and Differences to the NBTI/pMOS Case," in *Proc.IRPS*, pp. XT18.1–XT18.5, 2014.

[67] M. Waltl, G. Rzepa, A. Grill, W. Goes, J. Franco, B. Kaczer, L. Witters, J. Mitard, N. Horiguchi, and T. Grasser, "Superior nbti in high-k sige transistors — part ii: Experimental," *IEEE Trans.Elect.Dev.*, vol. 64, no. 5, pp. 2092–2098, 2017.

[68] M. Waltl, G. Rzepa, A. Grill, W. Goes, J. Franco, B. Kaczer, L. Witters, J. Mitard, N. Horiguchi, and T. Grasser, "Superior nbti in high-k sige transistors — part ii: Theory," *IEEE Trans.Elect.Dev.*, vol. 64, no. 5, pp. 2099–2105, 2017.

[69] A. Lelis and T. Oldham, "Time Dependence of Switching Oxide Traps," *IEEE Trans.Nucl.Sci.*, vol. 41, no. 6, pp. 1835–1843, 1994.

[70] E. Poindexter and W. Warren, "Paramagnetic Point Defects in Amorphous Thin Films of SiO_2 and Si_3N_4: Updates and Additions," *J.Electrochem.Soc.*, vol. 142, no. 7, pp. 2508–2516, 1995.

[71] T. Grasser, M. Waltl, Y. Wimmer, W. Gs, R. Kosik, G. Rzepa, H. Reisinger, G. Pobegen, A.-M. El-Sayed, A. Shluger, and B. Kaczer, "Gate-Sided Hydrogen Release as the Origin of "Permanent" NBTI Degradation: From Single Defects to Lifetimes," in *Proc.IEDM*, pp. 535–538, 2015.

[72] T. Grasser, M. Waltl, G. Rzepa, W. Goes, Y. Wimmer, A. M. El-Sayed, A. L. Shluger, H. Reisinger, and B. Kaczer, "The 'Permanent' Component of NBTI Revisited: Saturation, Degradation-Reversal, and Annealing," in *Proc.IRPS*, pp. 5A–2–1–5A–2–8, 2016.

[73] T. Aichinger, M. Nelhiebel, S. Einspieler, and T. Grasser, "*In Situ* Polyheater – A Reliable Tool for Performing Fast and Defined Temperature Switches on Chip," *IEEE Trans.Dev.Mater.Rel.*, vol. 10, no. 1, pp. 3–8, 2010.

[74] T. Grasser, W. Goes, Y. Wimmer, F. Schanovsky, G. Rzepa, M. Waltl, K. Rott, H. Reisinger, V. Afanas'ev, A. Stesmans, A.-M.El-Sayed, and A. Shluger, "On the Microscopic Structure of Hole Traps in pMOSFETs," in *Proc.IEDM*, pp. 21.1.1–21.1.4, 2014.

[75] H. Kupka, *Transitions in Molecular Systems*. Wiley-VCH, 2010. ISBN: 978-3-527-41013-2.

[76] M. Newton, "Quantum Chemical Probes of Electron-Transfer Kinetics: the Nature of Donor-Acceptor Interactions," *Chem. Rev.*, vol. 91, no. 5, pp. 767–792, 1991.

[77] F. Jensen, *Introduction to Computational Chemistry*. John Wiley & Sons, 1999.

[78] K. Mikkelsen and M. Ratner, "Electron Tunneling in Solid-State Electron-Transfer Reactions," *Chem. Rev.*, vol. 87, no. 1, pp. 113–153, 1987.

[79] P. Hänggi, P. Talkner, and M. Borkovec, "Reaction-Rate Theory: Fifty Years after Kramers," *Rev.Mod.Phys.*, vol. 62, no. 2, pp. 251–342, 1990.

[80] H. Oberhofer and J. Blumb, "Electronic Coupling Matrix Elements from Charge Constrained Density Functional Theory Calculations using a Plane Wave Basis Set," *Journ.Chem.Phys.*, vol. 133, p. 244105, 2010.

[81] B. Kaduk, T. Kowalczyk, and T. V. Voorhis, "Constrained Density Functional Theory," *Chem.Rev.*, vol. 112, no. 1, pp. 321–370, 2012. PMID: 22077560.

[82] J. Blumberger and K. McKenna, "Constrained density functional theory applied to electron tunnelling between defects in MgO," *Phys.Chem.Chem.Phys.*, vol. 15, pp. 2184–2196, 2013.

[83] K. McKenna and J. Blumberger, "Crossover from incoherent to coherent electron tunneling between defects in MgO," *Phys.Rev.B*, vol. 86, no. 24, p. 245110, 2012.

[84] Note that the term 'dynamical coupling' has been used in [41] but is also known under the term 'nonadiabatic' coupling [41, 76, 87]. Furthermore, it corresponds to the elements of the nonadiabacity operator [41].

[85] S. Gosavi and R. A. Marcus, "Nonadiabatic Electron Transfer at Metal Surfaces," *J.Phys.Chem.B*, vol. 104, no. 9, pp. 2067–2072, 2000.

[86] Following the nomenclature of Refs. [75, 87, 88] and numerous publications based on constrained DFT [80, 82, 83, 90, 91], this transformation defines the diabatic approximation throughout this work. It is noted that this term is frequently used for the static or crude Born-Oppenheimer approximation [75, 92, 93], which is a special variant of the diabatic approximation.

[87] G. Worth and L. Cederbaum, "Beyond Born-Oppenheimer: Molecular Dynamics Through a Conical Intersection," *Ann.Rev. of Phys.Chem.*, vol. 55, pp. 127–158, 2004.

[88] T. V. Voorhis, T. Kowalczyk, B. Kaduk, L.-P. Wang, C.-L. Cheng, and Q. Wu, "The diabatic picture of electron transfer, reaction barriers, and molecular dynamics," *Annual Review of Physical Chemistry*, vol. 61, pp. 149–170, 2010.

[89] Note that the term 'static' coupling is also known under the term 'electron/charge transfer integral' [76, 90, 91], or 'electronic matrix element' [80].

[90] A. Kubas, F. Hoffmann, A. Heck, H. Oberhofer, M. Elstner, and J. Blumberger, "Electronic Couplings for Molecular Charge Transfer: Benchmarking CDFT, FODFT, and FODFTB against High-Level Ab Initio Calculations," *Journal of Chemical Physics*, vol. 140, p. 104105, 2014.

[91] Q. Wu and T. V. Voorhis, "Extracting Electron Transfer Coupling Elements from Constrained Density Functional Theory," *Journ.Chem.Phys.*, vol. 125, p. 164105, 2006.

[92] R. Pässler, "Description of Nonradiative Multiphonon Transitions in the Static Coupling Scheme," *Czech.J.Phys.B*, vol. 24, pp. 322–339, 1974.

[93] A. Alkauskas, Q. Yan, and C. V. de Walle, "First-Principles Theory of Nonradiative Carrier Capture via Multiphonon Emission," *Phys.Rev.B*, vol. 90, no. 7, p. 075202, 2014.

[94] T. Markvart, "Semiclassical Theory of Non-Radiative Transitions," *Journ.Phys.C: Sol.State Phys.*, vol. 14, no. 29, p. L895, 1981.

[95] F. Schanovsky, O. Baumgartner, V. Sverdlov, and T. Grasser, "A Multi Scale Modeling Approach to Non-Radiative Multi Phonon Transitions at Oxide Defects in MOS Structures," *Journ. of Computational Electronics*, vol. 11, no. 3, pp. 218–224, 2012.

[96] K. Huang and A. Rhys, "Theory of Light Absorption and Non-Radiative Transitions in F-Centres," *Proceedings of the Royal Society of London. Series A*, vol. 204, pp. 406–423, 1950.

[97] E. Condon, "A Theory of Intensity Distribution in Band Systems," *Phys.Rev.*, vol. 28, no. 6, pp. 1182–1201, 1926.

[98] R. Berger, C. Fischer, and M. Klessinger, "Calculation of the Vibronic Fine Structure in Electronic Spectra at Higher Temperatures. 1. Benzene and Pyrazine," *Journ.Chem.Phys.A*, vol. 102, pp. 7157–7167, 1998.

[99] R. Borrelli and A. Peluso, "Perturbative Calculation of FranckCondon Integrals: New Hints for a Rational Implementation," *Journ*, vol. 129, p. 064116, 2008.

[100] R. Borrelli, A. Capobianco, and A. Peluso, "Franck-Condon Factors — Computational Approaches and Recent Developments," *Can.J.Chem.*, vol. 91, pp. 495–504, 2013.

[101] P.-A. Malmqvist and N. Forsberg, "Franck-Condon Factors for Multidimensional Harmonic Oscillators," *Chem.Phys.*, vol. 228, pp. 227–240, 1998.

[102] R. Marcus and N. Sutin, "Electron Transfer in Chemistry and Biology," *Biochemica et Biophysica*, vol. 811, pp. 265–322, 1985.

[103] P. Landsberg, *Recombination in Semiconductors*. Cambridge University Press, 2003. ISBN: 978052154343.

[104] A. Stoneham, "Non-Radiative Transitions in Semiconductors," *Rep.Prog.Phys.*, vol. 44, pp. 1251–1295, 1981.

[105] A. Alkauskas, J. Lyons, D. Steiauf, and C. V. de Walle, "First-Principles Calculations of Luminescence Spectrum Line Shapes for Defects in Semiconductors: The Example of GaN and ZnO," *Phys.Rev.Lett.*, vol. 109, no. 26, p. 267401, 2012.

[106] F. Schanovsky, W. Goes, and T. Grasser, "Multiphonon hole-trapping from first principles," *J.Vac.Sci.&Techn.B*, vol. 29, pp. 01A201–1, 2011.

[107] S. Ganichev, I. Yassievich, V. Perel, H. Ketterl, and W. Prettl, "Tunneling Ionization of Deep Centers in High-Frequency Electric Fields," *Phys.Rev.B*, vol. 65, p. 085203, 2002.

[108] T. Keil, "Shapes of Impurity Absorption Bands in Solids," *Phys.Rev.*, vol. 140, no. 2A, pp. A601–A617, 1965.

[109] C. Henry and D. Lang, "Nonradiative Capture and Recombination by Multiphonon Emission in GaAs and GaP," *Phys.Rev.B*, vol. 15, no. 2, pp. 989–1016, 1977.

[110] P. Schmidt, "Computationally Efficient Recurrence Relations for One-Dimensional Franck-Condon Overlap Integrals," *Molecular Physics*, vol. 108, no. 11, pp. 1513–1529, 2010.

[111] P. Schmidt, "Nonadiabatic Anharmonic Electron Transfer," *Journ.Chem.Phys.*, vol. 138, p. 124108, 2013.

[112] In this respect, the interested reader is referred to the section VI where the correctness of this assumption will be discussed in much greater detail.

[113] C. V. de Walle and J. Neugebauer, "First-Principles Calculations for Defects and Impurities: Applications to III-Nitrides," *J.Appl.Phys.*, vol. 95, no. 8, pp. 3851–3879, 2004.

[114] D. Drabold, *The Theory of Defects in Semiconductors*. 2007. ISBN-10 3-540-33400-9.

[115] C. Castleton, A. Höglund, and S. Mirbt, "Managing the Supercell Approximation for Charged Defects in Semiconductors: Finite-size Scaling, Charge Correction Factors, the Band-Gap Problem, and the *Ab-Initio* Dielectric Constant," *Phys.Rev.B*, vol. 73, no. 3, pp. 035215–035226, 2006.

[116] P. Blöchl, "First-Principles Calculations of Defects in Oxygen-Deficient Silica Exposed to Hydrogen," *Phys.Rev.B*, vol. 62, no. 10, pp. 6158–6179, 2000.

[117] B. Meinerzhagen, A. T. Pham, S. M. Hong, and C. Jungemann, "Solving boltzmann transport equation without monte-carlo algorithms - new methods for industrial tcad applications," in *2010 International Conference on Simulation of Semiconductor Processes and Devices*, pp. 293–296, 2010.

[118] O. Baumgartner, M. M. Karner, and H. Kosina, "Modeling of High-K Metal Gate Stacks Using the Non-Equilibrium Greens Function Formalism," in *Proc.SISPAD*, 2008.

[119] S. Datta, *Quantum Transport — Atom to Transistor*. Cambridge University Press, 2005.

[120] W. Goes, M. Toledano-Luque, O. Baumgartner, M. Bina, F. Schanovsky, B. Kaczer, and T. Grasser, "Understanding correlated drain and gate current fluctuations," in *Proc.IPFA*, pp. 51–56, July 2013.

[121] M. Karner, A. Gehring, S. Holzer, M. Pourfath, M. Wagner, W. Goes, M. Vasicek, O. Baumgartner, C. Kernstock, K. Schnass, G. Zeiler, T. Grasser, H. Kosina, and S. Selberherr, "A Multi-Purpose Schrödinger-Poisson Solver for TCAD Applications," *Journ. of Computational Electronics*, vol. 6, pp. 179–182, 2007.

[122] F. Schanovsky, O. Baumgartner, and T. Grasser, "Multi-Scale Modeling of Multi-Phonon Hole Capture in the Context of NBTI," in *Proc.SISPAD*, pp. 15–18, 2011.

[123] W. Goes, F. Schanovsky, and T. Grasser, *The Bias Temperature Instability: Experiment, Theory, and Modeling for Devices and Circuits* , ch. Advanced Modeling of Oxide Defects. Springer-Verlag, 2013.

[124] M. Lu and M. Chen, "Oxide-Trap-Enhanced Coulomb Energy in a Metal-Oxide-Semiconductor System," *Phys.Rev.B*, vol. 72, no. 23, pp. 235417–1, 2005.

[125] M. Schulz, "Coulomb Energy of Traps in Semiconductor Space-Charge Regions," *J.Appl.Phys.*, vol. 74, no. 4, pp. 2649–2657, 1993.

[126] O. Ibe, *Markov Processes for Stochastic Modeling*. Academic Press, 2009.

[127] J. Weeks, J. Tully, and L. Kimerling, "Theory of Recombination-Enhanced Defect Reactions in Semiconductors," *Phys.Rev.B*, vol. 12, no. 8, pp. 3286–3292, 1975.

[128] D. Gillespie, *Markov Processes: An Introduction for Physical Scientists*. Academic Press, 1992.

[129] T. Grasser, K. Rott, H. Reisinger, M. Waltl, J. Franco, and B. Kaczer, "A Unified Perspective of RTN and BTI," in *Proc.IRPS*, p. 4A.5.14A.5.7, 2014.

[130] C. Nicklaw, D. Fleetwood, R. Schrimpf, and S. Pantelides, "The Structure, Properties, and Dynamics of Oxygen Vacancies in Amorphous SiO_2," *IEEE Trans.Nucl.Sci.*, vol. 49, no. 6, pp. 2667–2673, 2002.

[131] A. Lelis, T. Oldham, H. Boesch, and F. McLean, "The Nature of the Trapped Hole Annealing Process," *IEEE*

Trans.Nucl.Sci., vol. 36, no. 6, pp. 1808–1815, 1989.

[132] F. Schanovsky, O. Baumgartner, W. Goes, and T. Grasser, "A Detailed Evaluation of Model Defects as Candidates for the Bias Temperature Instability," in *Proc.SISPAD*, pp. 1–4, 2013.

[133] J. Conley and P. Lenahan, "Molecular Hydrogen, E′ Center Hole Traps, and Radiation Induced Interface Traps in MOS Devices," *IEEE Trans.Nucl.Sci.*, vol. 40, no. 6, pp. 1335–1340, 1993.

[134] J. de Nijs, K. Druijf, V. Afanasev, E. van der Drift, and P. Balk, "Hydrogen-Induced Donor-Type Si/SiO_2 Interface States," *Appl.Phys.Lett.*, vol. 65, pp. 2428–2430, 1994.

[135] A.-M. El-Sayed, Y. Wimmer, W. Goes, T. Grasser, V. Afanas'ev, and A. Shluger, "Theoretical Models of Hydrogen-Induced Defects in Amorphous Silicon Dioxide," *Phys.Rev.B*, vol. 92, no. 1, pp. 014107–014118, 2015.

[136] A. El-Sayed, M. Watkins, T. Grasser, V. Afanas'ev, and A. Shluger, "Hydrogen-Induced Rupture of Strained SiO Bonds in Amorphous Silicon Dioxide," *Phys.Rev.Lett.*, vol. 114, no. 11, pp. 115503–115508, 2015.

[137] P. Bunson, M. D. Ventra, S. Pantelides, R. Schrimpf, and K. Galloway, "Ab Initio Calculations of H^+ Energetics in SiO_2: Implications for Transport," *IEEE Trans.Nucl.Sci.*, vol. 46, no. 6, pp. 1568–1573, 1999.

[138] A. Yokozawa and Y. Miyamoto, "First-Principles Calculations for Charged States of Hydrogen Atoms in SiO_2," *Phys.Rev.B*, vol. 55, no. 20, pp. 13783–13788, 1997.

[139] A. van Duin, A. Strachan, S. Stewman, Q. Zhang, X. Xu, and W. Goddard, "$ReaxFF_{SiO}$ Reactive Force Field for Silicon and Silicon Oxide Systems," *J.Phys.Chem.A*, vol. 170, no. 19, pp. 3803–3811, 2003.

[140] J. Fogarty, H. Aktulga, A. Grama, A. van Duin, and S. Pandit, "A Reactive Molecular Dynamics Simulation of the Silica-Water Interface," *Journ.Chem.Phys.*, vol. 132, no. 17, 2010.

[141] A.L. Al-Moatasem and M.B. Watkins and V. Afanas'ev and A.L. Shluger, "Nature of Intrinsic and Extrinsic Electron Trapping in SiO_2," *Phys.Rev.B*, vol. 89, no. 12, pp. 125201–125211, 2014.

[142] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, "Quickstep: Fast and Accurate Density Functional Calculations using a Mixed Gaussian and Plane Waves Approach ," *Computer Physics Communications*, vol. 167, no. 2, pp. 103 – 128, 2005.

[143] M. Guidon, J. Hutter, and J. VandeVondele, "Robust Periodic HartreeFock Exchange for Large-Scale Simulations Using Gaussian Basis Sets," *Journal of Chemical Theory and Computation*, vol. 5, no. 11, pp. 3010–3021, 2009. PMID: 26609981.

[144] K. Vollmayr, W. Kob, and K. Binder, "Cooling-Rate Effects in Amorphous Silica: A Computer-Simulation Study," *Phys.Rev.B*, vol. 54, no. 22, pp. 15808–15827, 1996.

[145] S. Susman, K. Volin, D. Price, M. Grimsditch, J. Rino, R. Kalia, P. Vashishta, G. Gwanmesia, Y. Wang, and R. Liebermann, "Intermediate-Range Order in Permanently Densified Vitreous SiO_2: A Neutron-Diffraction and Molecular-Dynamics study," *Phys.Rev.B*, vol. 43, no. 1, pp. 1194–1197, 1991.

[146] G. Henkelman, G. Jóhannesson, and H. Jnsson, "Methods for Finding Saddle Points and Minimum Energy Paths," in *Theoretical Methods in Condensed Phase Chemistry* (S. Schwartz, ed.), vol. 5 of *Progress in Theoretical Chemistry and Physics*, pp. 269–302, Springer Netherlands, 2002.

[147] V. Sulimov, P. Sushko, A. Edwards, A. Shluger, and A. Stoneham, "Asymmetry and Long-Range Character of Lattice Deformation by Neutral Oxygen Vacancy in α-Quartz," *Phys.Rev.B*, vol. 66, no. 2, p. 024108, 2002.

[148] P. Lenahan, "Deep Level Defects Involved in MOS Device Instabilities," *Microelectron.Reliab.*, vol. 47, pp. 890–898, 2007.

[149] J. Isoya, J. Weil, and L. Halliburton, "EPR and Ab-Initio SCFMO Studies of the $\mathrm{Si} \cdot \mathrm{HSi}$ system in the E'_4 center of αquartz," *Journ.Chem.Phys.*, vol. 74, no. 10, pp. 5436–5448, 1981.

[150] R. Weeks and C. Nelson, "Trapped Electrons in Irradiated Quartz and Silica: II, Electron Spin Resonance," *Journal of the American Ceramic Society*, vol. 43, no. 8, pp. 399–404, 1960.

[151] P. Bunson, M. D. Ventra, S. Pantelides, D. Fleetwood, and R. Schrimpf, "Hydrogen-Related Defects in Irradiated SiO_2," *IEEE Trans.Nucl.Sci.*, vol. 47, no. 6, pp. 2289–2296, 2000.

[152] W. Goes, M. Karner, V. Sverdlov, and T. Grasser, "Charging and Discharging of Oxide Defects in Reliability Issues," *IEEE Trans.Dev.Mater.Rel.*, vol. 8, no. 3, pp. 491–500, 2008.

[153] K.-O. Ng and D. Vanderbilt, "Structure and Oxidation Kinetics of the $\mathrm{Si(100)} - \mathrm{SiO_2}$ Interface," *Phys.Rev.B*, vol. 59, no. 15, pp. 10132–10137, 1999.

[154] M. Wilde and K. Fukutani, "Hydrogen Detection Near Surfaces and Shallow Interfaces with Resonant Nuclear Reaction Analysis," *Surface Science Reports*, vol. 69, no. 4, pp. 196–295, 2014.

[155] Z. Liu, S. Fujieda, H. Ishigaki, M. Wilde, and K. Fukutani, "Current Understanding of the Transport Behavior of Hydrogen Species in MOS Stacks and Their Relation to Reliability Degradation," *ECS Trans.*, vol. 35, no. 4, pp. 55–72, 2011.

[156] Y. Wimmer, A.-M. El-Sayed, W. Goes, T. Grasser, and A. Shluger, "Role of hydrogen in volatile behaviour of defects in SiO_2-based electronic devices," *Proc.Roy.Soc.A*, vol. 472, no. 2190, pp. 1–23, 2016.

[157] G. Rzepa, M. Waltl, W. Goes, B. Kaczer, and T. Grasser, "Microscopic Oxide Defects Causing BTI, RTN, and SILC on High-k FinFETs," in *Proc.SISPAD*, pp. 144–147, 2015.

[158] B. R. Tuttle, "Theoretical Investigation of the Valence-Band Offset between Si(001) and SiO_2," *Phys.Rev.B*, vol. 70, no. 12, p. 125322, 2004.

[159] P. Sushko, S. Mukhopadhyay, A. Stoneham, and A. Shluger, "Oxygen Vacancies in Amorphous Silica: Structure and Distribution of Properties," *Microelectron.Eng.*, vol. 80, pp. 292–295, 2005.

[160] A. Foster, V. Sulimov, F. L. Gejo, A. Shluger, and R. Nieminen, "Structure and Electrical Levels of Point Defects in Monoclinic Zirconia," *Phys.Rev.B*, vol. 64, no. 22, pp. 224108–224118, 2001.

[161] J. Gavartin, D. Mũnoz Ramo, A. Shluger, G. Bersuker, and B. Lee, "Negative Oxygen Vacancies in HfO_2 as Charge Traps in High-κ Stacks," *Appl.Phys.Lett.*, vol. 89, no. 8, p. 082908, 2006.

[162] D. Mũnoz Ramo, J. L. Gavartin, A. L. Shluger, and G. Bersuker, "Spectroscopic Properties of Oxygen Vacancies in Monoclinic HfO_2 Calculated with Periodic

and Embedded Cluster Density Functional Theory," *Phys.Rev.B*, vol. 75, no. 20, p. 205336, 2007.

[163] T. Grasser, K. Rott, H. Reisinger, M. Waltl, F. Schanovsky, and B. Kaczer, "NBTI in Nanoscale MOSFETs — The Ultimate Modeling Benchmark," *IEEE Elect.Dev.*, vol. 61, 2014.

[164] A. Grill, B. Stampfer, M. Waltl, K.-S. Im, J.-H. Lee, C. Ostermaier, H. Ceric, and T. Grasser, "Characterization and Modeling of Single Defects in GaN/AlGaN Fin-MIS-HEMTs," in *Proc.IRPS*, 2017.

[165] T. Grasser, "The Capture/Emission Time Map Approach to the Bias Temperature Instability," in *The Bias Temperature Instability: Experiment, Theory, and Modeling for Devices and Circuits*, T. Grasser (Ed.), Springer-Verlag, 2013.

[166] W. Nolting, *Grundkurs Theoretische Physik 5/2: Quantenmechanik - Methoden und Anwendungen.* Springer-Verlag, 2006.

[167] R. H. Fowler and L. Nordheim, "Electron Emission in Intense Electric Fields," in *Proc. R. Soc. Lond. A*, vol. 119, pp. 173–181, 1928.